National Library of Canada

Acquisitions and Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale du Canada

Direction des acquisitions et des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

Canada

UNIVERSITY OF ALBERTA

CRYSTALLOGRAPHIC ANALYSES OF A FLAVOPROTEIN INVOLVED IN BACTERIAL BIOLUMINESCENCE AND HUMAN PROGASTRICSIN, AN ASPARTIC PROTEINASE ZYMOGEN

by

STANLEY A. MOORE

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE

DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF BIOCHEMISTRY

EDMONTON, ALBERTA

FALL 1994

National Library
of Canada

Acquisitions and
Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale
du Canada

Direction des acquisitions et
des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

The author has granted an irrevocable non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

L'auteur a accordé une licence irrévocable et non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de sa thèse de quelque manière et sous quelque forme que ce soit pour mettre des exemplaires de cette thèse à la disposition des personnes intéressées.

The author retains ownership of the copyright in his/her thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without his/her permission.

L'auteur conserve la propriété du droit d'auteur qui protège sa thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

ISBN 0-315-95223-7

Canada

UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR:	Stanley A. Moore
TITLE OF THESIS:	Crystallographic Analyses of a Flavoprotein Involved in Bacterial Bioluminescence and Human Progastricsin, an Aspartic Proteinase Zymogen
DEGREE:	Doctor of Philosophy
YEAR THIS DEGREE GRANTED:	Fall 1994

Permission is hereby granted to the UNIVERSITY OF ALBERTA LIBRARY to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

Stanley A. Moore
Signature

20 Phillips Ave.
Permanent Address

Summerside P.E.I.

C1N 4L1

Date: Aug 26, 1994

UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled "Crystallographic Analyses of a Flavoprotein Involved in Bacterial Bioluminescence and Human Progastricsin, an Aspartic Protease Zymogen" submitted by Stanley A. Moore in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Michael James

Dr. Michael N.G. James

Randy Read

Dr. Randy J. Read

S E Jensen

Dr. Susan E. Jensen

A. R. Morgan

Dr. A. Richard Morgan

B W Matthews

Dr. Brian W. Matthews

Date: June 20, 1994

For my dad

William J. Moore

1925-1994

ABSTRACT

Bioluminescent bacteria belonging to the genus *Photobacterium* produce large quantities of a nonfluorescent flavoprotein (NFP) that is a homologue of the α and β subunits of luciferase. Single crystals of purified NFP have been grown from ammonium sulphate solutions. The crystals grow as thin (0.06 mm) plates and belong to the orthorhombic space group $C222_1$: a = 57.06 (3), b = 92.41 (6), c = 99.52 (6) Å. A native and two heavy atom derivative data sets have been collected. The three-dimensional structure of NFP was solved by isomorphous replacement.

The NFP polypeptide exists as a symmetrical homodimer. The monomer displays a novel fold that contains a seven-stranded β/α barrel. Two molecules of a myristylated flavin mononucleotide are associated with each NFP monomer. This adduct has C6 of the flavin isoalloxazine moiety covalently attached to the C3' methine carbon atom of the myristate. The chiral C3' atom is in the S-configuration in both adducts. Both myristyl groups are buried within the hydrophobic core of the protein.

The NFP atomic coordinates have been refined against synchrotron diffraction data extending to 1.6 Å resolution. The final R-factor is 0.175 for the 228 amino acid polypeptide, the two FMN adducts, one sulphate group, and 191 water molecules.

An analysis of the potential structural similarities between NFP and the bacterial luciferase subunits suggests that both luciferase subunits adopt a $(\beta\alpha)_8$ barrel fold. Amino acids conserved in NFP and luciferase cluster in three distinct regions of NFP which are located at hydrophobic interfaces between the β-strands and α-helices. A partial model of the luciferase β subunit from *P. leiognathi* has been built based on the refined structure of NFP.

In a separate analysis, the molecular structure of human progastricsin (hPGC) has been determined and refined to a conventional R-factor of 0.179 against data to 1.62 Å resolution. The 43 amino acid prosegment is similar in conformation to the prosegment of porcine pepsinogen (pPGN). However, there are large conformational differences centred around the segment containing Arg38p to Pro6, relative to the equivalent region in the structure of pPGN.

ACKNOWLEDGMENTS

This thesis is the result of two collaborative projects. The first involves bacterial bioluminescence. John Lee, Dennis O'Kane and Bruce Gibson from the University of Georgia were instrumental in supplying purified protein, encouragement, and much information on bacterial bioluminescence. I have really enjoyed my collaboration with John and Dennis and their work has laid down the cornerstone for Chapters 2 to 5. Bruce Gibson deserves recognition for purifying the protein used in this study.

For the human progastricsin work, I am indebted to Nadezhda Tarasova for supplying purified protein and Maia Chernaia for growing beautiful crystals and also preparing the heavy atom derivatives. Anita Sielecki did all of the PROLSQ and TNT refinement on human progastricsin. The painstaking care of her work is evident in the quality of the Ramachandran plot and the beautiful electron density maps in Chapter 6. Anita has been a sound source of guidance and advice, both personal and professional, during my stay in Edmonton. She also puts on a mean dinner party.

Mae Wylie has fearlessly taken on the task of typing up this thesis, including tedious sequence tables and other nasty bits. I cannot thank her enough for her patience and understanding, and undeniable skill. Thanks, Mae.

My intrepid supervisor, Mike James, has been supportive and a source of motivation during my stay in Edmonton. When things weren't working in the lab, he invariably found something else to keep my spirits up. Mike has a very keen eye for protein structures, and always finds that one hydrogen bond I leave out of figures. Thanks buddy. I'm not so sure many other supervisors could have understood my off-the-wall humor so well. Mike, and two other lab members, Natalie Strynadka and Marie Fraser, are responsible for the incredible NFP synchrotron data presented in Chapter 5. I owe them my life. Marie and Natalie have supplied much help and information over the years, on running programs and using that beast of an area detector in particular. They've also been good friends.

Randy Read, in the Medical Microbiology Department has always been accommodating in listening to my crystallographic and computer queries. Thanks Randy for taking the time. MIR will never be the same again. The James lab has been a good and productive place to work. Steve, Nina, Kathy, Ernst, Kui, and the rest of the gang, carry on dudes and good luck. Mark Israel, our programmer and resident opera singer

for many years, deserves mention. Mark made many easy-to-use and timesaving changes on such often-used programs as FRODO, MMS and HARDCOPY. Thanks, Mark. You've saved me untold hours of computer-related hardship.

The financial support of the Alberta Heritage Foundation for Medical Research, in the form of a Graduate Studentship to myself, has made the work contained in this thesis possible. I was also fortunate to be awarded a Dissertation Fellowship during my final year at the University of Alberta.

Many people in Edmonton made the place hospitable. Lee and Shirley Adams provided free food and lodging during my first six weeks in town. George and Ellen Strynadka have been terrific surrogate parents on more than one Christmas away from home. In the trout department, Darwyn Kobasa has been a willing accomplice on several flyfishing expeditions, and has always been there when it counted.

Finally, I want to thank my parents for letting me find my own way, even if it did take such a long time.

TABLE OF CONTENTS

Page

CHAPTER 1: INTRODUCTION 1

Bioluminescence 1

Bioluminescent bacteria 1

REFERENCES 7

CHAPTER 2: CRYSTALLIZATION OF *Photobacterium leiognathi* NFP, AN UNUSUAL FLAVOPROTEIN FROM A BIOLUMINESCENT BACTERIUM 12

INTRODUCTION 12

RESULTS 14

REFERENCES 17

CHAPTER 3: CRYSTAL STRUCTURE OF A FLAVOPROTEIN RELATED TO THE SUBUNITS OF BACTERIAL LUCIFERASE 21

INTRODUCTION 21

RESULTS 25

Crystallography 25

Molecular Structure 25

Dimer Interface 32

Flavin Adducts 32

DISCUSSION 35

Homology with Luciferase 35

MATERIALS AND METHODS 37

Structure Determination 37

Refinement 38

REFERENCES 39

Page

CHAPTER 4: COMMON STRUCTURAL FEATURES OF THE *luxF* PROTEIN AND THE SUBUNITS OF BACTERIAL LUCIFERASE: EVIDENCE FOR A $(\beta\alpha)_8$ FOLD IN LUCIFERASE 42

INTRODUCTION 42

MATERIALS AND METHODS 44

RESULTS 45

P. leiognathi luciferase β subunit model 45

Evidence for a $(\beta\alpha)_8$ structure 45

Similarities to and differences from glycolate oxidase 47

Conservation of NFP helices in luciferase 63

Conserved clusters involving surface loops 68

CONCLUSIONS 77

REFERENCES 78

CHAPTER 5: 1.6 Å RESOLUTION REFINEMENT OF THE NON-FLUORESCENT FLAVOPROTEIN FROM *Photobacterium leiognathi* 82

INTRODUCTION 82

MATERIALS AND METHODS 83

Data collection and processing 83

Least squares refinement 87

RESULTS 91

The β-Barrel 91

α-Helices 99

Reverse turns in the NFP structure 108

Asx turns in NFP 108

The flavin-fatty acid adducts 114

The dimer interface 114

Crystal packing 128

CONCLUSIONS 134

REFERENCES 135

Page

CHAPTER 6: CRYSTAL AND MOLECULAR STRUCTURES OF HUMAN PROGASTRICSIN AT 1.62 Å RESOLUTION 139

INTRODUCTION 139

MATERIALS AND METHODS 141

Crystallization 141

Data collection 144

Molecular replacement 144

Heavy atom derivatives 144

RESULTS 152

Overall description of the progastricsin molecule 152

Analysis of the prosegment 160

Electrostatic interactions in human progastricsin 175

The active site of human progastricsin 175

The activation process 178

REFERENCES 180

CHAPTER 7: DISCUSSION 187

NFP and bacterial bioluminescence 187

Human progastricsin 190

REFERENCES 192

LIST OF TABLES

		Page
Table 2.1	Crystallographic data for NFP from *P. leiognathi*	15
Table 3.1	Heavy atom data for NFP	27
Table 4.1	Structurally equivalent polypeptide segments in NFP and glycolate oxidase	51
Table 4.2	Barrel core residues in NFP and glycolate oxidase	56
Table 4.3	Helix start and end points in NFP and luciferase	65
Table 5.1	Photon factory imaging plate data for NFP	84
Table 5.2	Merging statistics for NFP data	85
Table 5.3	Completeness of NFP diffraction data as a function of resolution	86
Table 5.4	Refinement statistics for *P. leiognathi* NFP	88
Table 5.5	Regions of the NFP molecule with poorly associated electron density	90
Table 5.6	Hydrogen bonds associated with the β-sheet in NFP	96
Table 5.7	α-Helices in NFP	100
Table 5.8	i, i-5 Hydrogen bonds at the C-termini of α-helices	101
Table 5.9	Reverse (3^{10}) turns in the NFP structure	104
Table 5.10	Asx turns in the NFP structure	112
Table 5.11	Torsion angles for the FMN ribityl phosphate units	115
Table 5.12	Torsion angles for the myristic acid chains	116
Table 5.13	Hydrogen bonds at the dimer interface of NFP	122
Table 5.14	van der Waals contacts at the NFP dimer interface	124
Table 5.15	Protein-protein contacts in NFP involving crystallographic symmetry	130
Table 6.1a	Crystallographic data for human progastricsin	142
Table 6.1b	Heavy atom sites for progastricsin	145

Page

Table 6.2 Crystallographic refinement statistics for human progastricsin 148

Table 6.3 Disordered regions in the structure of human progastricsin 151

Table 6.4 Equivalent polypeptide segments in human progastricsin and porcine pepsinogen 155

Table 6.5 Protein-protein contacts in human progastricsin involving crystallographic symmetry 158

Table 6.6 Hydrogen bonds involving only main chain donor and acceptor atoms in the prosegment of human progastricsin 165

Table 6.7 Hydrogen bonds involving amino acid side chains in the prosegment of human progastricsin 167

Table 6.8 van der Waals contacts involving hydrophobic amino acid residues in the prosegment and mature enzyme regions of human progastricsin 170

Table 6.9 Charged residues in the prosegments and mature enzyme segments of human progastricsin and porcine pepsinogen 174

LIST OF FIGURES

		Page
Figure 1.1	Flavin compounds involved in bacterial bioluminescence	3
Figure 3.1a	Amino acid sequence alignment for NFP's and the luciferase β subunits	23
Figure 3.1b	Residues conserved between NFP and the luciferase subunits	24
Figure 3.2a	Schematic topology diagram of NFP	28
Figure 3.2b	Ribbon drawing of the NFP dimer	29
Figure 3.3	Electron density for the myristylated flavins at 2.3 Å resolution	31
Figure 3.4a	Protein-adduct interactions at the NFP dimer interface	34
Figure 3.4b	Protein-adduct interactions at the NFP C-terminus	34
Figure 4.1	Superposition of NFP and the partial luciferase β subunit model	46
Figure 4.2	Structural superposition of glycolate oxidase onto NFP	49
Figure 4.3	Residues making up the β-barrel core in NFP and glycolate oxidase	52
Figure 4.4	Ribbon diagrams of NFP and glycolate oxidase, highlighting common structural features	54
Figure 4.5	Secondary structure prediction for the luciferase domain not found in NFP	60
Figure 4.6	Residues conserved with luciferase near Trp96 in NFP	70
Figure 4.7	Conserved residues in the vicinity of helices α_2 and α_3	73
Figure 4.8	Conserved residues near Phe198 in NFP	76
Figure 5.1	A Ramachandran plot of the NFP model after least squares cycle 93	89
Figure 5.2	A Bhat OMIT map in the vicinity of Val74	92
Figure 5.3	Electron density for the flavin adduct at the dimer interface at 1.60 Å resolution	93
Figure 5.4	Electron density for the flavin adduct at the C-terminus	94

Page

Figure 5.5 Ribbon diagrams highlighting the β/α barrel in NFP 95

Figure 5.6 Hydrogen bonding in the NFP β-barrel 98

Figure 5.7 The C-capping box of helix α_7 102

Figure 5.8 A four-residue salt bridge on helix α_1 of NFP 103

Figure 5.9 Hydrogen bonds at the N-terminus of helix α_3 106

Figure 5.10 The kink in helix α_4 is mediated by water molecules 107

Figure 5.11 A combination of reverse and Asx turns centred at Asp58 109

Figure 5.12 The β-hairpin connecting strands β_3 and β_4 110

Figure 5.13 Hydrogen bonding between the side chains of reverse turns at the C-termini of strands β_2 and β_3 111

Figure 5.14 A Cys-containing Asx turn at Cys197 113

Figure 5.15 The myristic acid binding pocket at the dimer interface 117

Figure 5.16 A comparison of the conformations of the two flavin adducts..... 118

Figure 5.17 The solvent cavity at the NFP dimer interface 120

Figure 5.18 Hydrogen bonding between the NFP monomers is mediated by water molecules 121

Figure 5.19 The β_3-β_4 hairpin interacts with itself at the dimer interface 126

Figure 5.20 A side-on view of the dimer interface in NFP 127

Figure 5.21 A potential carbon-oxygen hydrogen bond at the dimer interface 129

Figure 5.22 Molecular packing in the NFP unit cell 131

Figure 5.23 Molecular packing in the ac plane in NFP 132

Figure 5.24 Molecular packing parallel to the crystallographic b axis 132

Figure 5.25 A sulphate anion makes bridging hydrogen bonds between histidines on symmetry-related helices of NFP 133

Figure 6.1 A Ramachandran plot of the refined human progastricsin model 147

Page

Figure 6.2 A $2|F_o|-|F_c|$ OMIT map of the peptide between residues Ala124 and Tyr125 149

Figure 6.3 A portion of the $3|F_o|-2|F_c|$ map at Arg322 150

Figure 6.4 A superposition of porcine pepsinogen onto human progastricsin 153

Figure 6.5 A superposition of porcine pepsin onto human progastricsin 154

Figure 6.6 A C^{α} drawing of hPGC, highlighting regions that differ in conformation with pPGN 156

Figure 6.7 Molecular packing in the hPGC unit cell 157

Figure 6.8 Specific symmetry contacts in the hPGC unit cell 159

Figure 6.9 A comparison of the prosegments of hPGC and pPGN 161

Figure 6.10 Amino acid sequence alignment for the prosegments of pepsinogens and progastricsins 163

Figure 6.11 Hydrogen bonding interactions in the prosegment of hPGC 166

Figure 6.12 Hydrophobic interactions between the prosegment and the mature enzyme 171

Figure 6.13 The four conserved salt bridges in hPGC 173

Figure 6.14 Conformational differences at the active sites of hPGC and pPGN 176

Figure 6.15 Hydrogen bonding interactions at the active site of hPGC 177

LIST OF ABBREVIATIONS AND SYMBOLS

Å	Ångstrom (10^{-10} metres)
A	Electrical current in amperes
ACP	Acyl carrier protein
AMP	Adenosine monophosphate
B	Isotropic thermal motion factor in $Å^2$ ($B=8\pi^2/3<r^2>$)
c	Velocity of electromagnetic radiation
cAMP	Cyclic adenosine monophosphate
D	Dalton (1/12 mass of ^{12}C nucleus)
DTT	Dithiothreitol
E	Energy
EDTA	Ethylenediamine tetraacetic acid
EMBL	European Molecular Biology Laboratory
EMP	Ethyl mercury phosphate
|F|	X-ray structure factor amplitude
FAD	Oxidized form of flavin adenine dinucleotide
FMN	Oxidized form of flavin mononucleotide
$FMNH_2$	Reduced form of flavin mononucleotide
FT*	Fluorescence transient
h	Planck's constant ($E = h\nu = hc/\lambda$)
hPGC	Human progastricsin (pepsinogen C)
hPGN	Human pepsinogen A
HPLC	High performance liquid chromatography
I	X-ray scattering intensity
λ	Wavelength of electromagnetic radiation
M.I.R.	Multiple isomorphous replacement
ν	Frequency of electromagnetic radiation
NFP	Nonfluorescent flavoprotein
nm	Nanometre (10^{-9} metres)
NMR	Nuclear magnetic resonance
pPGN	Porcine pepsinogen A
rms	Root mean square
σ	Standard deviation
V	Electrical potential in volts
V_m	Specific volume in $Å^3$/Dalton (also Matthew's coefficient)
W	Power in watts

CHAPTER 1

INTRODUCTION

Bioluminescence

Many marine organisms, and a few terrestrial ones, have the ability to produce their own light through biochemical reactions (Campbell, 1989; McCapra, 1990; Meighen, 1991). This phenomenon is generally termed bioluminescence and is found in a diverse collection of organisms including bacteria, fungi, algae, and many invertebrates. A well-known example of a bioluminescent terrestrial organism is the beetle *Photinus pyralis* (commonly called the firefly). Each group of organisms has evolved its own specialized system of generating and regulating the emission of light. Enzymes that catalyze light emission reactions are generally called luciferases and their luminescent substrates luciferins. The luciferins are a diverse group of compounds and include imidazolopyrazines (coelenterates and invertebrates), benzothiazoles (beetles). linear tetrapyrolles (dinoflagellates), flavins (fungi and bacteria) and aldehydes (several phyla) (Campbell, 1989; McCapra, 1990). The common denominator of this collection of compounds is that they all involve the decomposition of a peroxide intermediate to produce a singlet excited state that relaxes to the electronic ground state, liberating a photon of visible light (usually in the blue-green range). The luciferases that catalyze these luminescent chemical conversions are a diverse group of enzymes, and there is generally no phylogenetic relation between luciferases from different classes of organisms. The involvement of the peroxide is essential; it is one of the few chemical moieties that, upon decomposition, liberates enough energy to produce visible light, and does so in such a way as to produce products in electronically excited states (McCapra, 1990). Many of the more efficient bioluminescent reactions liberate CO_2 in the course of the peroxide decomposition. Another common feature of several bioluminescent organisms is that many transfer energy from the primary emitter to a secondary molecule that emits light at a different wavelength.

Bioluminescent bacteria

Bacteria capable of bioluminescence are generally marine in origin and are characteristically gram-negative motile rods. The most common genera of bioluminescent bacteria are *Photobacterium*, *Vibrio*, and *Alteromonas*, but there is also a terrestrial genus *Xenorhabdus* (Meighen, 1993; 1991). Unlike bioluminescence in many

of the eukaryotes, bacterial bioluminescence is continuous and not responsive to a mechanical stimulus. However many of the bioluminescent bacteria, especially the *Photobacteria*, are symbiotic with fishes (Haygood & Distel, 1993) and squids (Ruby & McFall-Ngai, 1992; McFall-Ngai & Ruby, 1991) in highly specialized light organs. Many of these specific bacterial symbionts are unculturable in the laboratory.

Bacterial bioluminescence is cell-density dependent, a threshold concentration of bacteria is required before light emission begins (reviewed in Meighen, 1991 and Hastings *et al.*, 1985). This phenomenon is known as autoinduction, and the concentration of a small, diffusable metabolite, the autoinducer, is responsible for the effect. The autoinducer has been identified as N-(β-ketocaproyl) homoserine lactone (Eberhard *et al.*, 1981) in *Vibrio fisheri*, and as N-(β-hydroxy butyryl) homoserine lactone in *Vibrio harveyi* (Cao & Meighen, 1989). Similar compounds that stimulate pathways involved in secondary metabolism have been identified in other groups of bacteria (Zhang *et al.*, 1993; Jones *et al.*, 1993). In *V. fisheri*, the *lux*I gene is responsible for synthesis of autoinducer. Bacterial bioluminescence is also under the control of catabolite repression, synthesis of the enzymes required for light emission being repressed in glucose-rich media and stimulated when the intracellular cAMP (cyclic adenosine monophosphate) level rises (reviewed in Meighen, 1993; 1991).

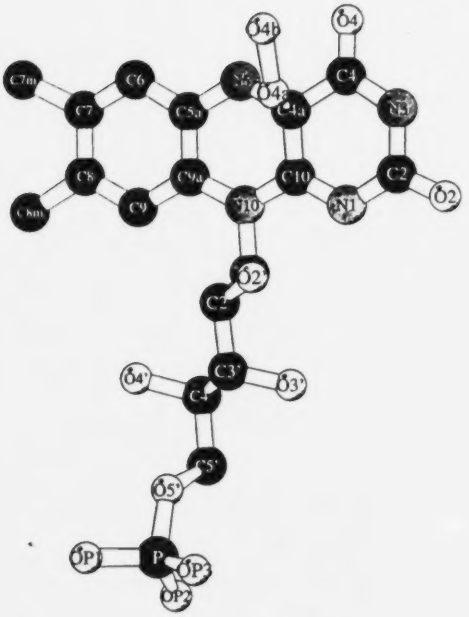
The genes responsible for bacterial bioluminescence are organized into an operon (the *lux* operon) that, in *Vibrio fisheri*, is under the transcriptional control of autoinducer. A repressor molecule (*luxR*) appears to be de-repressed when it binds to autoinducer. Recently, it has become apparent that regulation of the *lux* genes, at least in *V. harveyi*, is under the control of more than one signal transduction pathway, and is highly complex (Bassler & Silverman, 1993). At least two distinct autoinducers are now known to regulate bioluminescence in *V. fisheri* (Dunlap, 1993).

All bacterial *lux* operons sequenced to date contain five common genes. Two (*lux*A and B) code for the α and β subunits of bacterial luciferase respectively and are themselves evolutionarily related. The protein sequences of the α and β subunits are about 30% identical (Soly *et al.*, 1988). Bacterial luciferase is a flavin mono-oxygenase that requires $FMNH_2$ (reduced flavin mononucleotide), O_2, and myristic (C_{14}) aldehyde as substrates (reviewed in Baldwin & Ziegler, 1992). The enzyme is extremely slow, turning over once in ≈ 10s at room temperature under standard assay conditions. Most, if not all of the enzymatic activity resides on the α subunit. The β subunit seems to be

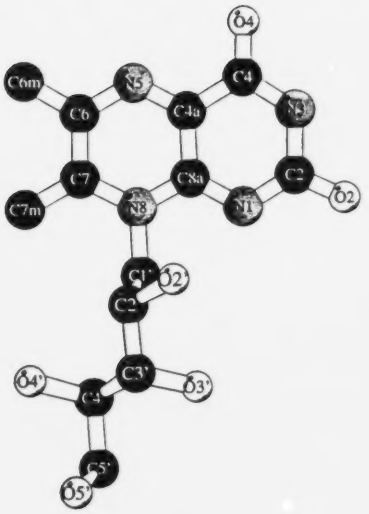
required for the correct folding and stability of the α subunit (Baldwin & Ziegler, 1992). Both subunits have a molecular weight of about 36 kD. The α subunit is slightly larger, having a highly-conserved 27 amino-acid insertion relative to the β subunit at position 260 of the *V. harveyi* α subunit. This region of the α subunit is susceptible to proteolytic digestion, but the presence of phosphate in the buffer is known to protect the enzyme against proteolysis. A conclusion consistent with these observations is that this region of the α subunit participates in the binding of the FMN phosphate at the luciferase active site (Baldwin & Ziegler, 1992).

All known enzyme-bound reduced flavins including luciferase-bound $FMNH_2$ are known to exist in the N1 anionic state (Vervoort *et al.*, 1986). $FMNH^-$ bound to luciferase rapidly reacts with oxygen to form a 4a-peroxy-4a,5-dihydroflavin (Hastings *et al.*, 1973). This intermediate is common to all known flavin monooxygenases (Ghishla & Massey, 1989; 1986) and its formation likely proceeds via a free radical mechanism in which the reduced flavin anion donates an electron (likely from N5 of the isoalloxazine) to triplet molecular oxygen. The resultant neutral flavin radical and the superoxide radical anion then undergo spin inversion and recombine to form the peroxide species. The chemical structure for the 4a-peroxy-4a,5-dihydroflavin involved in the bioluminescence reaction is given in Figure 1.1.

The luciferase 4a-peroxy-4a,5-dihydroflavin is unusually stable and can be isolated chromatographically at low temperatures (Hastings *et al.*, 1973). Upon aldehyde binding, a fluorescent transient species develops and slowly decays to the ground state, emitting visible light at 490 nm (reviewed in Lee *et al.*, 1991). The aldehyde likely forms a peroxyhemiacetal with the reduced flavin, and subsequently decomposes to form myristic acid and the excited singlet state of 4a-hydroxy-4a,5-dihydroflavin (Kurfürst *et al.*, 1984; Mancheroux *et al.*, 1993). Recent mechanistic studies have suggested that an important aspect of the reaction is donation of one electron from N5 of the isoalloxazine ring to the peroxide bond and subsequent bond scission to give the excited state of 4a-hydroxy-4a,5-dihydro-FMN (Eckstein *et al.*, 1993). Model 4a,5-dihydroflavins, substituted by electron donating or withdrawing substituents *para* to N5 at C8 of the isoalloxazine, give luciferase-catalyzed reaction rates that are correlated with the one-electron oxidation potentials of these compounds. The more readily the C8 substituent increases the electron density at N5 through mesomeric effects, the more rapidly the light emission step proceeds (Eckstein *et al.*, 1993). The final step in this proposed mechanism is the decomposition of the 4a-hydroxy-4a,5-dihydroflavin to give



4a-peroxy-4a,5-dihydroFMN



6,7-dimethyl-8-ribityllumazine

Figure 1.1 Two flavin compounds involved in the bacterial bioluminescence reaction: the 4a-peroxy-4a,5-dihydroflavin, which is the precursor to the excited-state primary emitter species, and 6,7-dimethyl-8-ribityllumazine, which is the ultimate energy acceptor and light emitter in *Photobacterium* species.

oxidized FMN (flavin mononucleotide) and water.

The myristic aldehyde substrate required for the bacterial luciferase reaction is derived from myristyl-ACP (Acyl Carrier Protein) in the bacterial fatty acid synthesis pathway. The other three genes (*lux*C, D and E) common to all bacterial *lux* operons are responsible for the shunting of myristyl-ACP from fatty acid synthetase and converting it to the myristic aldehyde (Meighen, 1993; 1991). The acyl transferase (product of *lux*D) is responsible for the specific transfer of myristate from myristyl-ACP (Ferri & Meighen, 1991) to the acyl protein synthetase and contains an active site serine residue (Soly & Meighen, 1991). The predicted amino acid sequence of the *lux*D gene indicates the acyl transferase is a homologue of serine esterases (Soly & Meighen, 1991). The transferase from *V. harveyi* has been crystallized (Swenson *et al.*, 1992) and its three-dimensional structure has been solved at 2.0 Å resolution (Z. Derewenda, personal communication); it has a structure similar to that of other serine esterases.

The next step in myristic aldehyde synthesis is catalyzed by acyl protein synthetase (product of *lux*E). It condenses the myristate with ATP to form myristyl-AMP, which is tightly bound to the enzyme (Meighen, 1991). This activated compound is then transferred to the fatty acid reductase (product of *lux*C) where a fatty acid thioester is formed upon release of AMP (Meighen, 1993). The thioester is then reduced by NADPH and hydrolyzed to yield the myristic aldehyde. The three *lux* enzymes involved in aldehyde formation in *Photobacterium phosphoreum* are assembled into a high molecular weight reductase complex (Meighen, 1993).

Several other genes genetically linked to the *lux* operons of bioluminescent bacteria appear to be involved with riboflavin synthesis, notably *lux*G and the *lux*H cluster (Lee & Meighen, 1992; Meighen, 1993; Richter *et al.*, 1992). In addition to these, antennae proteins are involved in altering the *in vivo* bioluminescence spectrum in certain *Vibrio fisheri* strains (*lux*Y) and many *Photobacterium* isolates (*lux*L). These antennae proteins both appear to be derived from the α subunit of bacterial riboflavin synthetase (O'Kane & Prasher, 1992; O'Kane *et al.*, 1991). The *lux*Y protein binds riboflavin and shifts the *in vivo* bioluminescence spectrum to yellow wavelengths (Daubner *et al.*, 1987), whereas the *lux*L protein binds 6,7-dimethyl-8-(1'-D-ribityl) lumazine and shifts the bioluminescence spectrum to shorter wavelengths (Koka & Lee, 1979; O'Kane *et al.*, 1985; Lee *et al.*, 1991).The structure of the lumazine chromophore is given in Figure 1.1.

Another gene, *lux*F, is found between *lux*B and *lux*E in *Photobacterium phosphoreum* (Mancini *et al.*, 1988) and has been cloned and sequenced. The predicted polypeptide product is 231 amino acids in length and is a homologue of the bacterial luciferase subunits (Soly *et al.*, 1988). The same gene has also been cloned and sequenced in *P. leiognathi* (Baldwin *et al.*, 1989; Illarionov *et al.*, 1990). The predicted polypeptide is 228 amino acids in length and is about 50% identical to *lux*F from *P. phosphoreum*. The predicted amino acid sequence of *lux*F from *P. leiognathi* is identical with the experimentally-determined N-terminal amino-acid sequence of a non-fluorescent flavoprotein (NFP) isolated from *P. leiognathi* (O'Kane *et al.*, 1987; Raibekas, 1991). An identical result has been found in *P. phosphoreum* (Kasai *et al.*, 1987; 1991). The NFP's are co-induced with luciferase in a cell-density-dependent manner and are produced in large quantities by these bacteria. The protein binds an unusual flavin chromophore that is a covalent adduct of myristate and FMN (Kasai *et al.*, 1982; 1991). The function of NFP is not presently known. The majority of this thesis concerns the crystallization, structure determination, and high resolution refinement of the NFP from *P. leiognathi*. The relationship of NFP to the subunits of bacterial luciferase is discussed in detail.

REFERENCES

Baldwin, T. O. & Ziegler, M. M. (1992). The biochemistry and molecular biology of bacterial bioluminescence. In *Chemistry and Biochemistry of Flavoenzymes* (Müller, F., ed.), vol. III, pp. 467-530, CRC, Boca Raton, FL.

Baldwin, T. O., Devine, J. H., Heckel, R. C., Lin, J. W. & Shadel, G. S. (1989). The complete nucleotide sequence of the *lux* regulon of *Vibrio fisheri* and the *lux* ABN region of *Photobacterium leiognathi* and the mechanism of control of bacterial bioluminescence. *J. Biol. Chemilumin.* **4**, 326-341.

Bassler, B. L. & Silverman, M. R. (1993). Intercellular signalling in *Vibrio harveyi*: Regulation of expression of bioluminescence. *Bioluminescence Symposium*, Maui, Hawaii. November 5-10, 1993. Abstr. Sp35.

Campbell, A. K. (1989). Living light: Biochemistry, function and biomedical applications. In *Essays in Biochemistry* (Marshall, R. D. & Tipton, K. F., eds.), vol. 24, pp. 41-81, The Biochemical Society by Academic Press, New York.

Cao, J.-G. & Meighen, E. A. (1989). Purification and structural identification of an autoinducer for the luminescence system of *Vibrio harveyi*. *J. Biol. Chem.* **264**, 21670-21676.

Daubner, S. C., Astorga, A. M., Leisman, G. B. & Baldwin, T. O. (1987). Yellow light emission of *Vibrio fisheri* strain Y1: Purification and characterization of the energy accepting yellow fluorescent protein. *Proc. Natl. Acad. Sci. USA* **84**, 8912-8916.

Dunlap, P. V. (1993). A genetic approach to understanding the regulation and function of the *Vibrio fisheri* luminescence system. *Bioluminescence Symposium*, Maui, Hawaii. November 5-10, 1993. Abstr. Sp30.

Eberhard, A., Burlingame, A. L., Eberhard, C., Kenyon, G. L., Nealson, K. H. & Oppenheimer, N. J. (1981). Structural identification of autoinducer of *Photobacterium fischeri* luciferase. *Biochemistry* **20**, 2444-2449.

Eckstein, J. W., Hastings, J. W. & Ghisla, S. (1993). Mechanism of bacterial bioluminescence: 4a,5-dihydroflavin analogs as models for luciferase hydroperoxide intermediates and the effect of substituents at the 8-position of flavin on luciferase kinetics. *Biochemistry* **32**, 404-411.

Ferri, S. R. & Meighen, E. A. (1991). A *lux*-specific myristoyl transferase in luminescent bacteria related to eukaryotic serine esterases. *J. Biol. Chem.* **266**, 12852-12857.

Ghisla, S. & Massey, V. (1986). New flavins for old: artificial flavins as active site probes of flavoproteins. *Biochem. J.* **239**, 1-12.

Ghisla, S. & Massey, V. (1989). Mechanisms of flavoprotein-catalyzed reactions. *Eur. J. Biochem.* **181**, 1-17.

Hastings, J. W., Balny, C., Peuch, C. L. & Douzou, P. (1973). Spectral properties of an oxygenated luciferase - flavin intermediate isolated by low-temperature chromatography. *Proc. Natl. Acad. Sci. USA* **70**, 3468-3472.

Hastings, J. W., Potrikus, C. J., Gupta, S. C., Kurfürst, M. & Makemson, J. C. (1985). Biochemistry and physiology of bioluminescent bacteria. In *Advances in Microbial Physiology* (Rose, A. H. & Tempest, D. W., eds.), vol. 26, pp. 235-291, Academic Press, New York.

Haygood, M. G. & Distel, D. L. (1993). Bioluminescent symbionts of flashlight fishes and deep-sea anglerfishes form unique lineages related to the genus *Vibrio*. *Nature* **363**, 154-156.

Illarionov, B. A., Blinov, V. M., Donchenko, A. P., Protopopova, M. V., Karginov, V. A., Mertvetsov, N. P. & Gitelson, J. I. (1990). Isolation of bioluminescent functions from *Photobacterium leiognathi*: analysis of *lux*A, *lux*B, *lux*G and neighboring genes. *Gene* **86**, 89-94.

Jones, S., Yu, B., Bainton, N. J., Birdsall, M., Bycroft, B. W., Chhabra, S. R., Cox, A. J. R., Golby, P., Reeves, P. J., Stephens, S., Winson, M. K., Salmond, G. P. C., Stewart, G. S. A.B. & Williams, P. (1993). *The lux* autoinducer regulates the production of exoenzyme virulence determinants in *Erwinia carotovora* and *Pseudomonas aeruginosa*. *EMBO J.* **12**, 2477-2482.

Kasai, S., Fujii, S., Miura, R., Odani, S., Nakaya, T. & Matsui, K. (1991). Structure of FP390 including its prosthetic group (Q-flavin):physiological significance of light emitting reaction in luminous bacteria. In *Flavins and Flavoproteins* (Curti, B., Ronchi, S. & Zanetti, G., eds.), pp. 285-288. Walter de Gruyter & Co., Berlin.

Kasai, S., Matsui, K. & Nakamura, T. (1982). Structure of P-Flavin from *P. phosphoreum*. In *Flavins and Flavoproteins* (Massey, M. & Williams, C. H., eds.), pp. 459-462, Elsevier, North Holland.

Kasai, S., Matsui, K. & Nakamura, T. (1987). Purification and some properties of FP390 from *P. phosphoreum*. In *Flavins and Flavoproteins* (Edmondson, D. E. & McCormick, D. B., eds.), pp. 647-650, Walter de Gruyter & Co., Berlin.

Koka, P. & Lee, J. (1979). Separation and structure of the prosthetic group of the blue fluorescence protein from the bioluminescent bacterium *Photobacterium phosphoreum*. *Proc. Natl. Acad. Sci. USA* **76**:3068-3072.

Kurfürst, M., Ghisla, S. & Hastings, J. W. (1984). Characterization and postulated structure of the primary emitter in the bacterial luciferase reaction. *Proc. Natl. Acad. Sci. USA* **81**, 2990-2994.

Lee, C. Y. & Meighen, E. A. (1992). The *lux* genes in *Photobacterium leiognathi* are closely linked with genes corresponding in sequence to riboflavin synthesis genes. *Biochem. Biophys. Res. Commun.* **186**, 690-697.

Lee, J., Matheson, I. B. C., Müller, F., O'Kane, D. J., Vervoort, J. & Visser, A. J. (1991). The mechanism of bacterial bioluminescence. In *Chemistry and Biochemistry of Flavoenzymes* (Müller, F., ed.), pp. 109-151, CRC, Boca Raton, FL.

Macheroux, P., Ghisla, S. & Hastings, J. W. (1993). Spectral detection of an intermediate preceding the excited state in the bacterial luciferase reaction. *Biochemistry* **32**, 14183-14186.

Mancini, J. A., Boylan, M., Soly, R., Graham, A. F. & Meighen, E. A. (1988). Cloning and expression of the *Photobacterium phosphoreum* luminescence system demonstrates a unique *lux* gene organization. *J. Biol. Chem.* **263**, 14308-14314.

McCapra, F. (1990). The chemistry of bioluminescence: Origins and mechanism. In *Light and Life in the Sea* (Herring, P. J., Campbell, A. K., Whitfield, M. & Maddock, L., eds.), Chapter 17, pp. 265-278, Cambridge University Press, Cambridge.

McFall-Ngai, M. J. & Ruby, E. G. (1991). Symbiont recognition and subsequent morphogenesis as early events in an animal-bacterial mutualism. *Science* **254**, 1491-1494.

Meighen, E. A. (1991). Molecular biology of bacterial bioluminescence. *Microbiol. Rev.* **55**, 123-142.

Meighen, E. A. (1993). Bacterial bioluminescence: organization, regulation, and application of the *lux* genes. *FASEB J.* **7**, 1016-1022.

O'Kane, D. J. & Prasher, D. C. (1992). Evolutionary origins of bacterial bioluminescence. *Mol. Microbiol.* **6**, 443-449.

O'Kane, D. J., Karle, V. A. & Lee, J. (1985). Purification of lumazine proteins from *Photobacterium leiognathi* and *Photobacterium phosphoreum*. Bioluminescence properties. *Biochemistry* **24**, 1461-1467.

O'Kane, D. J., Vervoort, J., Muller, F. & Lee, J. (1987). Purification and characterization of an unusual nonfluorescent flavoprotein from *Photobacterium leiognathi*. In *Flavins and Flavoproteins* (Edmondson, D. E. & McCormick, D. B., eds.), pp. 641-645, Walter de Gruyter & Co., Berlin.

O'Kane, D. J., Woodward, B., Lee, J. & Prasher, D. C. (1991). Borrowed proteins in bacterial bioluminescence. *Proc. Natl. Acad. Sci.* **88**, 1100-1104.

Raibekas, A. A. (1991). Green flavoprotein from *P. leiognathi*: purification, characterization and identification as the product of the *lux* G(N) gene. *J. Biolumin. Chemilumin.* **6**, 169-176.

Richter, G., Volk, R., Krieger, C., Lahm, H. W., Rothlisberger, U. & Bacher, A. (1992). Biosynthesis of riboflavin: cloning, sequencing and expression of the gene coding for 3, 4-dihydroxy-2-butanone 4-phosphate synthase of *Escherichia coli*. *J. Bacteriol.* **174**, 4050-4056.

Ruby, E. G. & McFall-Ngai, M. J. (1992). A squid that glows in the night: Development of an animal-bacterial mutualism. *J. Bact.* **174**, 4865-4870.

Soly, R. R. & Meighen, E. A. (1991). Identification of the acyl transfer site of fatty acyl-protein synthetase from bioluminescent bacteria. *J. Mol. Biol.* **219**, 69-77.

Soly, R. R., Mancini, J. A., Ferri, S. R., Boylan, M. & Meighen, E. A. (1988). A new *lux* gene in bioluminescent bacteria codes for a protein homologous to the bacterial luciferase subunits. *Biochem. Biophys. Res. Commun.* **155**, 351-358.

Swenson, L. Ferri , S. R., Green, R., Sharp, A. M., Meighen, E. A. & Derewenda, Z. S. (1992). Expression, purification and crystallization of the *Vibrio harveyi* acyltransferase. *J. Mol. Biol.* **227**, 572-574.

Vervoort, J., Muller, F., O'Kane, D. J., Lee, J. & Bacher, A. (1986). Bacterial luciferase: A carbon-13, nitrogen-15, and phosphorus-31 nuclear magnetic resonance investigation. *Biochemistry* **25**, 8067-8075.

Zhang, L., Murphy, P. J., Kerr, A. & Tate, M. E. (1993). *Agrobacterium* conjugation and gene regulation by *N*-acyl-L-homoserine lactones. *Nature* **362**, 446-450.

CHAPTER 2

CRYSTALLIZATION OF *PHOTOBACTERIUM LEIOGNATHI* NFP, AN UNUSUAL FLAVOPROTEIN FROM A BIOLUMINESCENT BACTERIUM[1]

INTRODUCTION

Bioluminescence in marine bacteria results from the following enzyme-catalyzed sequence of reactions in which myristic(C_{14}) aldehyde reacts with a luciferase-bound 4a-peroxy-4a,5-dihydroflavin. The resulting excited state fluorescent transient species (FT*) liberates a photon at λ = 490 nm and decays to flavin mononucleotide (FMN), H_2O and the corresponding fatty acid (see Lee *et al.*, 1991 and Ghisla *et al.*, 1987 for reviews of the reaction mechanism).

$$FMNH^- + O_2 + H^+ \rightarrow FMNHOOH$$
$$FMNHOOH + RCHO \rightarrow \rightarrow FT^*$$
$$FT^* \rightarrow FMN + H_2O + RCOOH + h\upsilon \quad \lambda = 490\ nm$$

The bioluminescence activity of marine bacteria belonging to the genera *Vibrio* and *Photobacterium* is conferred by the *lux* operon (Meighen, 1991), a group of genes that codes for five common gene products. Three of these, *lux* C, D, and E code for enzymes required to synthesize the myristic aldehyde substrate of the bioluminescence reaction whereas *lux* A and B code for the α and β subunits, respectively, of the light emitting enzyme bacterial luciferase, a heterodimeric protein of mw 76,000. Luciferase genes from four marine bacteria: *Vibrio harveyi* (Cohn *et al.*, 1985; Johnston *et al.*, 1986), *Vibrio fischeri* (Foran & Brown, 1988; Baldwin *et al.*, 1989), *Photobacterium leiognathi* (Illarionov *et al.*, 1990), *Photobacterium phosphoreum* (Meighen, 1991), one unidentified marine species isolated from the flashlight fish *Kryptophanaron alfredi* (Haygood, 1990), and the terrestrial bacterium *Xenorhabdus luminescens* (Johnston *et al.*, 1990; Szittner & Meighen, 1990), have been sequenced and the deduced amino-acid sequences show roughly 30% amino acid identity between the α and β subunits.

The α subunit sequences are more strongly conserved (54% to 88% pairwise identity) than the β subunit sequences (45% to 77% pairwise identity) amongst the

---

[1] A version of this chapter has been published. Moore *et al.*, 1992, *Journal of Molecular Biology* **224**:523-526.

marine bacteria (Meighen, 1991). This is consistent with genetic evidence that maps lesions affecting catalytic activity to the α subunit (Cline & Hastings, 1972). It is also consistent with subunit modification studies (Meighen *et al.*, 1971) in which only reconstituted heterodimers with unmodified α subunits have close to wild-type enzymatic activity. Temperature sensitive mutants map predominantly to the β subunit and it is believed that the β subunit plays an essential structural role in the enzyme (Cline & Hastings, 1972). Hybrid luciferases have also been constructed and have kinetic properties characteristic of the parent α subunit (Ruby & Hastings, 1980; Meighen & Bartlett, 1980), although Meighen & Bartlett observed that initial $FMNH_2$ binding was more closely related to the source of the β subunit. There is also some NMR evidence for a second, weak, binding site for $FMNH_2$ on *V. harveyi* luciferase (Vervoort *et al.*, 1986) and recently it has been shown that both the individual α and β subunits have small but genuine bioluminescence activity (Waddle & Baldwin, 1991). Clearly, structural information pertaining to the β subunit could have relevance to how luciferase binds the flavin and aldehyde substrates. Crystals of *V. harveyi* luciferase have previously been reported (Swanson *et al.*, 1985).

The *lux* operons of most *Photobacterium* species have an additional gene, *lux* F, inserted between *lux* B and *lux* E (Illarionov *et al.*, 1988; Mancini *et al.*, 1988; Baldwin *et al.*, 1989; Illarionov *et al.*, 1990) which codes for a 227 to 231 amino acid polypeptide of 26,000 Da (Soly *et al.*, 1988; Illarionov *et al.*, 1990) that has a high degree of sequence identity to the carboxy-terminal two thirds of the luciferase enzyme subunits. When aligned, the lux F amino-acid sequences from *P. leiognathi* and *P. phosphoreum* are 54% identical; they also have from 25% to 30% identity to the luciferase β subunit amino-acid sequences and have from 13% to 19% identity to the α subunit amino-acid sequences. A protein from *Photobacterium leiognathi* that copurifies with luciferase (O'Kane *et al.*, 1986) has a characteristic flavin-type coherent anti-Stokes Raman scattering spectrum (Visser *et al.*, 1983) and an N-terminal amino-acid sequence identical to that of the *lux* F gene product (O'Kane, unpublished observations; Raibekas, 1991). This protein binds an unusual flavin chromophore which contains FMN and myristate. The chromophore can only be removed from the protein by drastic chemical methods (O'Kane *et al.*, 1987; Raibekas, 1991) and a structure for the chromophore from *P. phosphoreum* has been proposed (Kasai *et al.*, 1991). The holoprotein is nonfluorescent in solution and hence has been called the nonfluorescent flavoprotein (NFP) (O'Kane *et al.*, 1987). NFP has also been isolated from *Photobacterium phosphoreum* and it is

produced concomitantly with luciferase in a cell density dependent manner (Kasai *et al.*, 1987). In spite of this, production of light by *P. phosphoreum lux* genes in *E. coli* does not appear to require *lux* F (Meighen, 1991).

The role of NFP in *Photobacterium* species is unclear but it is interesting to note that flavin chromophores with similar spectroscopic and chromatographic properties to the NFP chromophore have been isolated from luciferase preparations from *P. phosphoreum* (Matsuda & Nakamura, 1972; Kasai *et al.*, 1982) and *V. harveyi* (Mitchell & Hastings, 1970).

RESULTS

Here we report the crystallization of NFP from *P. leiognathi* and the characterization of the crystals by X-ray diffraction methods. Recently, crystallization conditions and preliminary X-ray characterization of the NFP homologue (FP_{390}) from *Photobacterium phosphoreum* have been reported (Kita *et al.*, 1991). NFP was isolated from strain S1 of *P. leiognathi* by sequential hydrophobic interaction, ion exchange and size exclusion HPLC similar to a previously reported protocol for purifying luciferase (O'Kane *et al.*, 1986). Purified NFP was stored in 25% v/v glycerol, 50 mM Na/K phosphate pH 7.0, 1 mM EDTA and 5 mM DTT at -60°C until required. Protein samples for crystallization were dialysed exhaustively against 10 mM Na citrate pH 6.0 in darkness and concentrated using Amicon Centricon tubes. The final protein concentration for crystallization was 5-10 mg/ml based on amino acid analysis and a mw of 26,897 for NFP (including chromophore) using the published amino-acid sequence for *P. leiognathi* NFP, strain Pl 741 (Baldwin *et al.*, 1989). Crystals were grown by the hanging-drop vapour-diffusion technique using equal volumes of protein solution and precipitant solution in the drop. The best precipitant was 30-34% saturated ammonium sulphate at pH = 5.5 (50 mM Na citrate buffer) with 3% v/v isopropanol added to reduce aggregation of the resulting thin plates. Large single crystals could only be grown using a rabbit whisker to seed pre-equilibrated drops with nuclei from previous crystals or by washing and reseeding small single crystals produced by the rabbit whisker technique. The largest single crystals were of maximum dimensions 0.70 mm along *a*, 0.35 mm along *c* and 0.06 mm along *b*. Screened X-ray precession photographs were recorded on an Enraf Nonius precession camera using graphite monochromated $CuK\bar{\alpha}$ radiation from a Rigaku RU-200 rotating anode X-ray generator operating at 40 kV, 150 mA (6kW). The symmetry and conditions limiting Bragg reflection on hk0 and h0l

Table 2.1

Crystallographic data for NFP from P. leiognathi

Space group	$C222_1$ (#20 International Tables for X-ray Crystallography)
Resolution Limit	2.2 Å
Unit Cell Dimensions:	
a	57.06(3) Å
b	92.41(6) Å
c	99.52(6) Å
Solvent Content (1 monomer per a.u.)	50.7%
V_m	2.44 $Å^3$/D
No. of Reflections between ∞ and 2.5 Å:	
Collected	9356, 99.5% complete
†Observed	8806, 93.7% complete
R_{merge} [a](Merging Agreement Index) = 0.044	

$$^{a}R_{merge} = \frac{\sum_{hkl} \sum_{i} |I_{i,hkl} - \langle I \rangle_{hkl}|}{\sum_{hkl} \sum_{i} I_{i,hkl}}$$

†Observed criteria: $|F| > 3\sigma|F|$, average redundancy of measurement per reflection is 7-fold.

diffraction patterns determined the space group unambiguously as $C222_1$ (International Tables #20) with unit cell dimensions $a = 57.06(3)$, $b = 92.41(6)$, $c = 99.52(6)$ Å. A complete native intensity data set to 2.5 Å resolution has been collected on twin San Diego multiwire detectors (Xuong *et al.*, 1985) using graphite monochromated $CuK\bar{\alpha}$ radiation at 6 kW from a Rigaku RU-200 rotating anode source. Data beyond 2.5 Å are relatively weak owing to the thinness of the crystals although there were reflections with $I > 5\sigma I$ at spacings < 2.2 Å resolution. Refined unit cell parameters and data processing statistics are given in Table 2.1. Given that $V_m = 2.44$ Å^3/Dalton (51% solvent) for NFP crystals (Matthews, 1968) it is most likely that there is one NFP monomer per crystallographic asymmetric unit and since NFP is a stable dimer in solution (O'Kane *et al.*, 1987; Raibekas, 1991) it is reasonable to conclude that the NFP dimer must be located on one of the crystallographic twofold axes along (x,0,0) or (0,y,1/4).

Determination of the NFP structure by crystallographic methods will facilitate unambiguous identification of the NFP chromophore and should explain how the protein binds so tenaciously to its ligand. It may also be possible to map approximately the flavin and aldehyde binding regions of bacterial luciferase, if the chromophore binding site of NFP is coincident with amino acid sequences that are conserved in the luciferase subunits. This in turn would open the way to rational design of site specific mutants to probe luciferase structure and function.

REFERENCES

Baldwin, T. O., Devine, J. H., Heckel, R. C., Lin, J.-W. & Shadel, G. S. (1989). The complete nucleotide sequence of the *lux* regulon of *Vibrio fischeri* and the *lux*ABN region of *Photobacterium leiognathi* and the mechanism of control of bacterial bioluminescence. *J. Biolumin. Chemilumin.* **4**, 326-341.

Cline, T. W. & Hastings, J. W. (1972). Mutationally altered bacterial luciferase. Implications for subunit functions. *Biochemistry* **11**, 3359-3370.

Cohn, D. H., Mileham, A. J., Simon, M. I., Nealson, K. H., Rausch, S. K., Bonham, D. & Baldwin, T. O. (1985). Nucleotide sequence of the *lux*A gene of *Vibrio harveyi* and the complete amino acid sequence of the α subunit of bacterial luciferase. *J. Biol. Chem.* **260**, 6139-6146.

Foran, D. R. & Brown, W. M. (1988). Nucleotide sequence of the *lux*A and *lux*B genes of the bioluminescent marine bacterium *Vibrio fischeri*. *Nucl. Acids Res.* **16**, 777.

Ghisla, S., Eckstein, J. & Macheroux, P. (1987). On the chemical mechanisms of bacterial luciferase, an updating. In *Flavins and Flavoproteins* (Edmondson, D. E. & McCormick, D. B., eds.), pp. 601-612, Walter de Gruyter & Co., Berlin.

Haygood, M. G. (1990). Relationship of the luminous bacterial symbiont of the Carribean flashlight fish *Kryptophanaron alfredi* (family Anomalopidae), to other luminous bacteria based on bacterial luciferase (*lux* A) genes. *Arch. Microbiol.* **154**, 496-503.

Illarionov, B. A., Blinov, V. M., Donchenko, A. P., Protopopova, M. V., Karginov, V. A., Mertvetsov, N. P. & Gitelson, J. I. (1990). Isolation of bioluminescent functions from *Photobacterium leiognathi*: analysis of *lux*A, *lux*B, *lux*G and neighbouring genes. *Gene* **86**, 89-94.

Illarionov, B. A., Protopopova, M. V., Karginov, V. A., Mertvetsov, N. P. & Gitelson, J. I. (1988). Nucleotide sequence of part of *Photobacterium leiognathi lux* region. *Nucl. Acids Res.* **16**, 9855.

International Tables for Crystallography. Volume A: Space Group Symmetry. (1983). (Hahn, T., ed.), D. Reidel Publishing, Boston.

Johnston, T. C., Rucker, E. B., Cochrum, L., Hruska, K. S. & Vandergraft, V. (1990). The nucleotide sequence of *lux*A and *lux*B genes of *Xenorhabdus luminescens* HM and a comparison of the amino acid sequences of luciferases from four species of bioluminescent bacteria. *Biochem. Biophys. Res. Commun.* **170**, 407-415.

Johnston, T. C., Thompson, R. B. & Baldwin, T. O. (1986). Nucleotide sequence of the *lux*B gene of *Vibrio harveyi* and the complete amino acid sequence of the β subunit of bacterial luciferase. *J. Biol. Chem.* **261**, 4805-4811.

Kasai, S., Fujii, S., Miura, R., Odani, S., Nakaya, T. & Matsui, K. (1991). Structure of FP390 including its prosthetic group (Q-flavin):physiological significance of light emitting reaction in luminous bacteria. In *Flavins and Flavoproteins* (Curti, B., Ronchi, S. & Zanetti, G., eds.), pp. 285-288. Walter de Gruyter & Co., Berlin.

Kasai, S., Matsui, K. & Nakamura, T. (1982). Structure of P-Flavin from *P. phosphoreum.* In *Flavins and Flavoproteins* (Massey, M. & Williams, C. H., eds.), pp. 459-462, Elsevier, North Holland.

Kasai, S., Matsui, K. & Nakamura, T. (1987). Purification and some properties of FP390 from *P. phosphoreum.* In *Flavins and Flavoproteins* (Edmondson, D. E. & McCormick, D. B., eds.), pp. 647-650, Walter de Gruyter & Co., Berlin.

Kita, A., Kasai, N., Kasai, S., Nakaya, T. & Miki, K. (1991). Crystallization and preliminary X-ray diffraction studies of a flavoprotein, FP_{390}, from a luminescent bacterium, *Photobacterium phosphoreum. J. Biochem. (Tokyo)* **110**, 748-750.

Lee, J., Matheson, I. B. C., Muller, F., O'Kane, D. J., Vervoort, J. & Visser, A. J. W. G. (1991). The mechanism of bacterial bioluminescence. In *Chemistry and Biochemistry of Flavoenzymes* (Muller, F., ed.), pp. 109-151, C. R. C. Press, Boca Raton, FA.

Mancini, J. A., Boylan, M., Soly, R., Graham, A. F. & Meighen, E. A. (1988). Cloning and expression of the *Photobacterium phosphoreum* luminescence system demonstrates a unique *lux* gene organization. *J. Biol. Chem.* **263**, 14308-14314.

Matsuda, M. & Nakamura, T. (1972). Isolation and partial characterisation of an enzyme-bound pigment. *J. Biochem.* **72**, 951-955.

Matthews, B. W. (1968). Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491-497.

Meighen, E. A. & Bartlett, I. (1980). Complementation of subunits from different bacterial luciferases. *J. Biol. Chem.* **265**, 11181-11187.

Meighen, E. A. (1991). Molecular biology of bacterial bioluminescence. *Microbiol. Rev.* **55**, 123-142.

Meighen, E. A., Zeigler-Nicoli, M. & Hastings, J. W. (1971). Functional differences of the nonidentical subunits of bacterial luciferase. Properties of hybrids of native and chemically modified bacterial luciferase. *Biochemistry* **10**, 4069-4073.

Mitchell, G. W. & Hastings, J. W. (1970). Light-induced bioluminescence. Isolation and characterization of a specific protein involved in the absorption and delayed emission of light. *Biochemistry* **9**, 2699-2707.

O'Kane, D. J., Ahmad, M., Matheson, I. B. C. & Lee, J. (1986). Purification of bacterial luciferase by high-performance liquid chromatography. In *Methods Enzymol.* (DeLuca, M. A. & McElroy, W. D., eds.), **133**, 109-128.

O'Kane, D. J., Vervoort, J., Muller, F. & Lee, J. (1987). Purification and characterization of an unusual nonfluorescent flavoprotein from *Photobacterium leiognathi*. In *Flavins and Flavoproteins* (Edmondson, D. E. & McCormick, D. B., eds.), pp. 641-645, Walter de Gruyter & Co., Berlin.

Raibekas, A. A. (1991). Green flavoprotein from *P. leiognathi*: purification, characterization and identification as the product of the *lux* G(N) gene. *J. Biolumin. Chemilumin.* **6**, 169-176.

Ruby, E. G. & Hastings, J. W. (1980). Formation of hybrid luciferases from subunits of different species of *Photobacterium*. *Biochemistry* **19**, 4989-4993.

Soly, R. R., Mancini, J. A., Ferri, S. R., Boylan, M. & Meighen, E. A. (1988). A new *lux* gene in bioluminescent bacteria codes for a protein homologous to the bacterial luciferase subunits. *Biochem. Biophys. Res. Commun.* **155**, 351-358.

Swanson, S., Weaver, L. H. , Remington, S. J., Matthews, B. W. & Baldwin, T. O. (1985). Crystals of luciferase from *Vibrio harveyi*, a preliminary characterization. *J. Biol. Chem.* **260**, 1287-1289.

Szittner, R. & Meighen, E. (1990). Nucleotide sequence, expression, and properties of luciferase coded by *lux* genes from a terrestrial bacterium. *J. Biol. Chem.* **265**, 16581-16587.

Vervoort, J., Muller, F., O'Kane, D. J., Lee, J. & Bacher, A. (1986). Bacterial luciferase: A carbon-13, nitrogen-15, and phosphorous-31 nuclear magnetic resonance investigation. *Biochemistry* **25**, 8067-8075.

Visser, A. J. W. G., Vervoort, J., O'Kane, D. J., Lee, J. & Carreira, L. A. (1983). Raman spectra of flavin bound in flavodoxins and in other flavoproteins. *Eur. J. Biochem.* **131**, 639-645.

Waddle, J. & Baldwin, T. O. (1991). Individual α and β subunits of bacterial luciferase exhibit bioluminescence activity. *Biochem. Biophys. Res. Commun.* **178**, 1188-1193.

Xuong, N. H., Nielson, C., Hamlin, R. & Anderson, D. (1985). Strategy for data collection from protein crystals using a multiwire counter area detector diffractometer. *J. Appl. Cryst.* **18**, 342-350.

CHAPTER 3

CRYSTAL STRUCTURE OF A FLAVOPROTEIN RELATED TO THE SUBUNITS OF BACTERIAL LUCIFERASE[1]

INTRODUCTION

Bioluminescent bacteria are the most abundant of the luminous organisms and occupy a wide variety of habitats in marine and terrestrial environments (Meighen, 1991; Hastings *et al.*, 1985). The production of light by these organisms is the result of an enzyme-catalyzed oxidation of myristic aldehyde to myristate by the flavin monooxygenase bacterial luciferase (Baldwin & Ziegler, 1991; Lee *et al.*, 1991). The reaction obeys the following stoichiometry: $FMNH_2 + O_2 + RCHO \rightarrow FMN + H_2O + RCOOH + h\upsilon$. The quantum yield of emitted light is about 10% for the purified enzyme and the emission maximum is approximately 490 nm (Lee *et al.*, 1991). The *in vivo* bioluminescence spectrum can be quite different in some luminous bacteria, due to the participation of antenna proteins, notably lumazine protein (Lee *et al.*, 1991; O'Kane *et al.*, 1985) and yellow fluorescence protein (Daubner *et al.*, 1987). The bioluminescence phenotype is conferred by the *lux* operon which contains the genes required for fatty aldehyde synthesis (*lux* C, D, E), and the luciferase subunits (*lux* A and B). Luciferase is a tightly associated $\alpha\beta$ heterodimer (Baldwin & Ziegler, 1992) of mw 76 kD, and the subunits exhibit 30% amino acid identity when their sequences are aligned (Soly *et al.*, 1988). Another gene, *lux* F, has been found only in species of *Photobacterium* (Meighen, 1991). It codes for a 228-231 amino acid nonfluorescent flavoprotein (NFP) of unknown function that is coinduced with luciferase (Kasai *et al.*, 1987). NFP has also been isolated and purified (O'Kane *et al.*, 1987). The *lux*F protein contains a tightly bound chromophore that is a covalent adduct of two of the products of the bioluminescence reaction: myristate and FMN (Kasai *et al.*, 1991). We have chosen to study the *lux*F protein, as it displays considerable amino-acid sequence homology with the luciferase subunits (Soly *et al.*, 1988; O'Kane & Prasher, 1992) (Figure 3.1a).

---

[1]A version of this chapter has been published. Moore *et al.*, 1993, *The EMBO Journal* **12**:1767-1774. However the legend to Figure 3b in that paper contains a mistake and should read (3'-S-myristyl) instead of 3'-R-myristyl.

Figure 3.1 (a) An alignment of the sequences of NFPs from *P. leiognathi* (Illarionov *et al.*, 1990) and *P. phosphoreum* (Soly *et al.*, 1988) and luciferase β subunits from *P. leiognathi* (Baldwin *et al.*, 1981) and *P. phosphoreum* (Meighen, 1991) is presented. The sequences for *P. leiognathi* NFP and the *P. phosphoreum* luciferase β subunits are numbered for convenience. The alignment uses all of the currently available luciferase and NFP protein sequences (Meighen, 1991; Soly *et al.*, 1988; Baldwin *et al.*, 1989; Illarionov *et al.*, 1990) and is similar to other published alignments (Soly *et al.*, 1988; O'Kane & Prasher, 1992). Insertions and gaps in the alignment have been positioned by paying close attention to the structure of NFP and are restricted to surface loops in NFP. A 102-103 amino acid insertion present in the luciferase subunits maps to the hairpin joining strands β_3 and β_4 (see Figure 3.2a) and so would be expected to be located at the N-terminal spacing of the β-barrel, i.e. on the opposite end of the barrel to the site of helices α_4 and α_5. Another 27 amino acid insertion present only in the luciferase α subunit is indicated and maps to the region between helices α_4 and α_5 at the C-terminal opening of the β-barrel. There are 15 (*) absolutely conserved residues in all published luciferase and NFP sequences and an additional 16 (†) that are strongly conserved. There are also 16 (‡) residues absolutely conserved between the NFP's and the luciferase β subunits. The majority of these conserved groups of amino acids cluster between residues 76 and 153 of *P. leiognathi* NFP, corresponding to helix α_2, strand β_5, helix α_3, strand β_6 and helix α_4. Strand β_1 also seems remarkably well conserved in the luciferase β subunits.

Figure 3.1a

P. leiognathi luxF
P. phosphoreum luxF
P. leiognathi luxB
P. phosphoreum luxB
102-103 amino acid insertion in both luciferase subunits
27 amino acid insertion in luciferase α-subunit

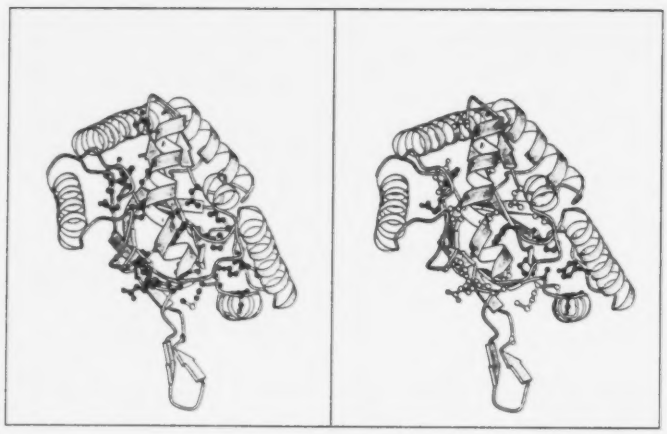


Figure 3.1 (b) A stereo MOLSCRIPT diagram (Kraulis, 1991) of the NFP monomer showing residues likely conserved in the luciferase subunits. The fifteen absolutely conserved amino acid side chains are shown (Figure 3.1a) in dark grey, including Gly7, Leu68, Gly69, Ala76, Ser78, Ala86, Pro91, Trp96, Tyr110, Ala114, His126, Ala141, Gly172, Phe198 and Glu199. Strongly conserved side chains are shown in light grey. The flavin cofactors are not shown to preserve clarity.

Secondary structural units were calculated by DSSP (Kabsch & Sander, 1983); β_1: Tyr6-Leu11, α_1: Pro21-Glu37, β_2: Val45-Ser48, β_{2a}: Lys54-Tyr56, β_{2b}: Thr60-Leu62, β_3: Phe65-Leu68, β_4: Gln71-Leu75, α_2: Pro79-Ala86, β_5: Leu92-Phe94, α_3: Gln100-Phe117, β_6: His126-Val134, α_4: Pro138-Tyr155, α_5: Ile162-Asn167, β_{6a}: Ala170-Gly172, α_6: Phe174-Gly187, β_7: Val192-Cys197, α_7: Gln204-Glu224.

RESULTS

Crystallography

The structure of NFP was solved using isomorphous replacement methods augmented with anomalous scattering information (Table 3.1), and solvent flattening procedures. The molecule is a symmetric dimer and sits on the 0,y,1/4 two-fold axis in space group $C222_1$ (Figure 3.2b). Each protein monomer in the asymmetric unit also binds two molecules of a myristylated FMN (Figures 3.3, 3.4). All amino acid residues are clearly visible in the refined electron density map. The current model includes the 228 amino-acids of the polypeptide, two molecules of myristylated FMN, and 102 water molecules. The R-factor is 17.8% for 10514 reflections with $|F| \leq 2\sigma |F|$ between 10 and 2.3 Å (87.6% complete). The overall geometry is good (0.013 Å r.m.s. distances, 0.031 Å r.m.s. angle related distances, 0.007 Å r.m.s. planar groups) and the B-factors* are generally low. The average B-factor for all main chain atoms is 21.0 $Å^2$ and 23.1 $Å^2$ for all side chain atoms. There is only one intermolecular contact in the direction of the crystallographic b-axis which suggests why the plate-like crystals used in this study were so thin (0.06 mm in the [010] direction) (Moore *et al.*, 1992).

Molecular Structure

NFP is a tightly associated dimer (O'Kane *et al.*, 1987; Raibekas, 1991) (Figure 3.2). The monomer consists of an open seven-stranded, mostly parallel β-barrel (strands β_3 and β_4 form an antiparallel hairpin) with seven α helices facing the solvent (Figures 3.1b and 3.2). There is a considerable gap between strands β_2 and β_3 at the dimer interface and this gap is bridged by a long β-meander (Ile52 to Ala63) that protrudes from the molecular surface. This meander contains a type I turn which has two additional residues forming antiparallel β-sheet hydrogen bonds and is labeled β_{2a} and β_{2b} in Figure 3.2a. The β_{2a}-β_{2b} hairpin follows and is an intimate part of the dimer interface. Including strand β_4, there are four more βα units except that helices α_4 and α_5 interrupt the topology between strand β_6 and helix α_6. Helix α_4 is kinked at residue 149, presumably to accommodate the packing of tyrosine residues in the hydrophobic core at positions 148 and 152. This helix forms part of the rim of the barrel's C-terminus. Helix α_5 is short and packs perpendicularly to helices α_4 and α_6. At the end

---

* $B = 8\pi^2/3 \langle r^2 \rangle$

Table 3.1

$|F_P|$, $|F_{PH}|$, and $|F_H|$ are the structure factor amplitudes for the native protein crystal, the heavy atom derivative crystal, and the calculated heavy atom substructure for each derivative crystal respectively.

$\sum_{hkl}$ indicates summation over hkl indices in reciprocal space.

$\sum_{i}$ indicates summation over i measurements of a particular *hkl* intensity.

$\langle I \rangle_{hkl}$ indicates the average of i measurements of a particular *hkl* intensity.

RMS = root mean square.

$$R_{merge} = \frac{\sum_{hkl} \sum_{i} |I_{i,hkl} - \langle I \rangle_{hkl}|}{\sum_{hkl} \sum_{i} |I_{i,hkl}|}$$

$$R_{iso} = \frac{\sum_{hkl} ||F_P| - |F_{PH}||}{\sum_{hkl} |F_P|}$$

$$\text{Phasing Power} = \left(\frac{\sum_{hkl} |F_H|^2}{\sum_{hkl} (|F_{PH}(obs)| - |F_{PH}(calc)|)^2} \right)^{1/2}$$

$$\text{R factor} = \frac{\sum_{hkl} ||F|_{obs} - |F_{calc}|}{\sum_{hkl} |F|_{obs}}$$

Table 3.1

Heavy atom data for NFP

	Unit cell parameters, (Å)	Unique Reflections (∞ 2.4 Å)	R_{merge}	Heavy atom soaking conditions	R_{iso}	Heavy atom parameters	Phasing power	Mean FOM*
Native data	a = 57.03 b = 92.25 c = 99.44	10199 (96.3% complete)	0.044 (7-fold redundant)					0.54
Ethyl Mercury Phosphate	a = 57.04 b = 92.47 c = 99.41	9754 (92% complete)	0.046 (3-fold redundant)	2 mM, 1 day†	0.178	x = 0.2320 y = 0.7613 z = 0.0211 B = 19.5 Å^2 Occ = 67.0 elec. Anom occ = 6.1 elec.	2.49	
Methyl Mercury Chloride	a = 57.31 b = 92.93 c = 99.88	8282 (78% complete)	0.083 (3-fold redundant)	1 mM, 6 hrs†	0.205	x = 0.2326 y = 0.7615 z = 0.0219 B = 15.1 Å^2 Occ = 63.0 elec. Anom occ = 7.1 elec.	1.60	

* Reflections from 15-2.4 Å resolution.

† Heavy atom solutions were made up in 45% saturated ammonium sulphate, 50 mM sodium citrate at pH 5.5.

Figure 3.2 (a) A schematic topology diagram of NFP showing the major secondary structural units.

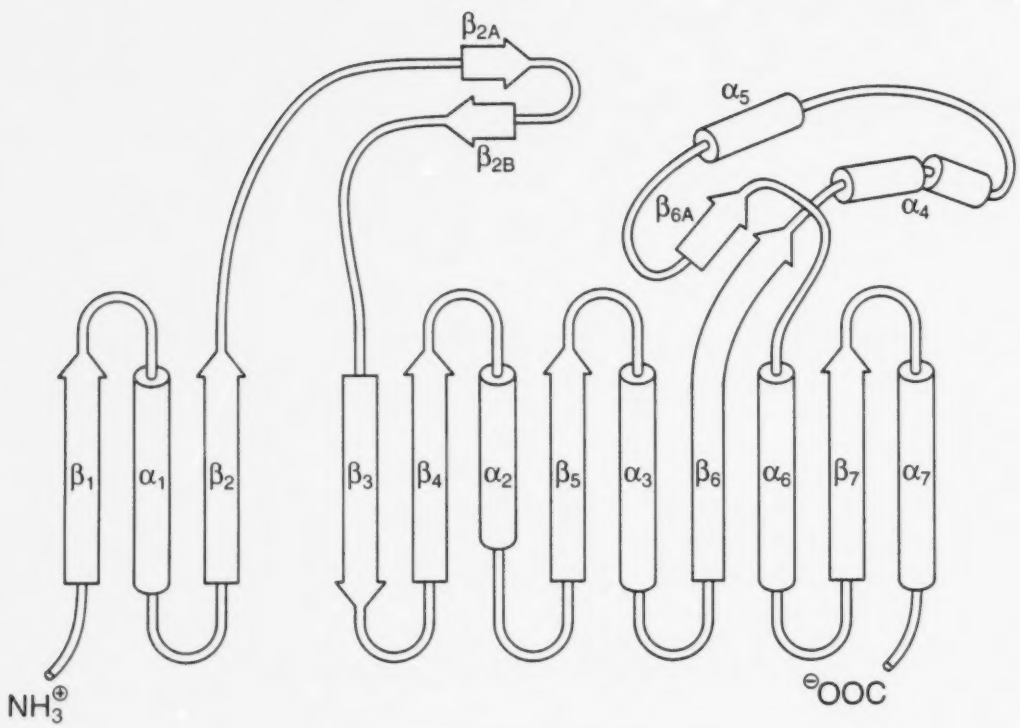


Figure 3.2 (b) A stereo MOLSCRIPT diagram (Kraulis, 1991) of the NFP dimer looking approximately down the crystallographic twofold axis at 0,y,1/4. The flavin myristates are represented as dark ball and stick models. Each monomer has approximately 46% α helix and 22% β strand (secondary structure calculated according to Kabsch & Sander, 1983).

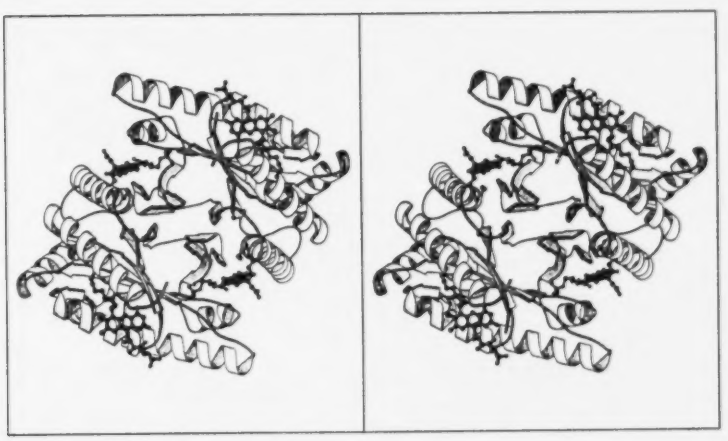


Figure 3.3 (a) Electron density for the flavin myristate at the dimer interface.

Figure 3.3 (b) Electron density for the flavin myristate at the C-terminus. The maps are contoured at the r.m.s. electron density of the unit cell (0.29e/Å^3) using coefficients $2m|F_o|-D|F_c|$ (Read, 1986). The flavin cofactor is modeled as 6-(3'-S-myristyl)-FMN using model FMN (Karplus & Schulz, 1987) and fatty acid coordinates (A.R. Sielecki, personal communication). The fatty acyl chain adopts a curved conformation in both FMN adducts. The C6 FMN-C3' fatty acid bond length was restrained to 1.50 Å and the bond angles at C3' of the fatty acid were made tetrahedral. During refinement, planar restraints were imposed on the isoalloxazine ring system. The origin of the myristylated flavin is a mystery although the C6 of isoalloxazine is a known site of covalent attachment to some flavoproteins (Singer & McIntire, 1984) and 6-hydroxy flavins have been characterized (Mayhew *et al.*, 1974). It has been suggested that compounds similar to the myristylated FMN could be suicide inhibitors of luciferase (Hastings *et al.*, 1985), but the chemistry of their possible formation during the bioluminescence reaction is not presently clear.

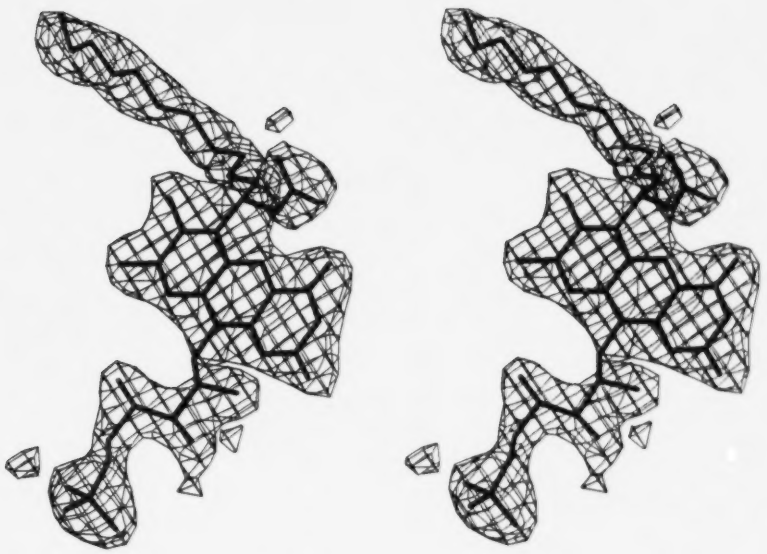


Figure 3.3a

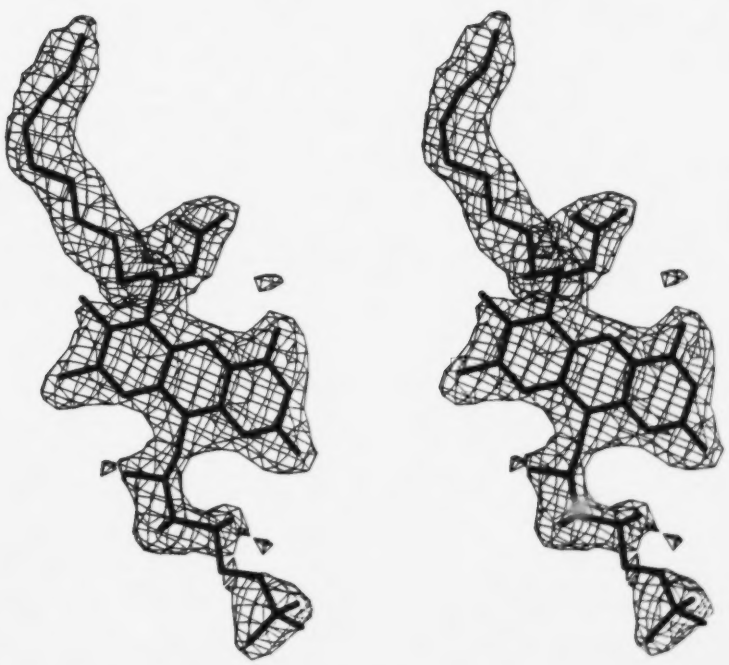


Figure 3.3b

of helix α_5, a stretch of three residues (strand β_{6a}) which is not part of the barrel makes parallel β-sheet hydrogen bonds to strand β_6. Strand β_7 forms β-sheet hydrogen bonds with strand β_1 thus completing the barrel. The C-terminal helix α_7 packs in a parallel fashion against helix α_1.

Dimer Interface

The dimer interface is made up of a relatively flat surface of each monomer that includes residues from the segment of polypeptide from Val8 to Ile72 (β_1 to β_4). One of the flavin myristates is involved in dimer contacts as well (Figures 3.2b and 3.4a). Each monomer contributes twenty-three side-chain to side-chain or side-chain to main-chain hydrogen bonds to this interaction. There are twelve hydrophobic side chains on each monomer that make van der Waals interactions (≤ 4.0 Å) at the interface. There are no main chain β-sheet hydrogen-bonding interactions between the monomers. Upon dimerization, approximately 27% (3200 $Å^2$) of the solvent-accessible surface area of each monomer is buried at the interface (calculated by the method of Connolly, 1983). This is one of the largest reported buried surface areas for a protein dimer (Argos, 1988). The myristylated flavin cofactor at the interface contributes approximately 20% of the buried solvent accessible surface area. Polar amino-acid side chains from helix α_2 interact with the ribityl group of the flavin and the carboxylate of the myristic acid. The fatty acid acyl chain is sandwiched between hydrophobic side chains from strands β_3 and β_4 and from the symmetry related strands β_1', β_2' and helix α_1'. A large channel extends from the molecular surface to a cavity at the interface. This cavity is lined with well ordered solvent molecules and also contains an intramolecular salt bridge. The water molecules in this cavity are generally hydrogen bonded to each other and to the main chain amide groups of amino acid residues 45 to 47, 49 to 55 and 67 to 69.

Flavin Adducts

There is a second myristylated FMN bound in the vicinity of the N and C termini of the molecule (Figures 3.2b and 3.4). Both cofactor sites have fewer hydrogen bonds to the FMN moiety than do other FMN and FAD binding proteins (Watt *et al.*, 1991; Lindqvist & Brändén, 1989; Karplus & Schulz, 1987). Approximately 80% (215 $Å^2$)of the solvent-accessible surface area of the isoalloxazine nucleus is buried at both sites. At the dimer interface mostly the pyrimidine ring is solvent accessible, whereas the benzenoid ring of the isoalloxazine is solvent exposed at the C-terminal site. The acyl

Figure 3.4 (a) Protein-adduct interactions at the NFP dimer interface. Interactions from the dyad related monomer are indicated with a prime. Hydrogen bonds are marked with dashed lines. The binding site is made up of secondary structure units β_3, β_4, α_2 and $\beta_{1'}$, $\alpha_{1'}$, $\beta_{2'}$. Most of the hydrophobic interactions come from the $\beta_{1'}$ $\alpha_{1'}$ $\beta_{2'}$ unit whereas most of the polar interactions are contributed from residues on helix α_2. The ribityl phosphate group makes hydrogen bonds with the O^{γ} oxygens of Thr77, Ser78 and Thr81 and also forms a salt bridge with Lys61. A water molecule also makes a hydrogen bond to one of the phosphate oxygens. Lys84, Thr81, and a water molecule make hydrogen bonds to the ribityl hydroxyl groups. The side chains of Lys87 and Tyr88 make hydrogen bonds to the myristyl carboxylate. Both N1 and O2 of the isoalloxazine pyrimidine ring act as hydrogen bond acceptors for an ordered water molecule. This water also accepts a hydrogen bond from the amide nitrogen of Lys61. Tyr14' stacks against the isoalloxazine pyrimidine ring and its hydroxyl group makes a hydrogen bond to the carbonyl oxygen of Leu62 prior to the beginning of strand β_3. The side chain amide of Asn12' also makes a hydrogen bond across the dimer interface to the carbonyl oxygen of Ala63.

Figure 3.4 (b) Protein-adduct interactions at the NFP C-terminus. Only atoms N3 and O4 of the isoalloxazine and the myristyl carboxylate make direct hydrogen bonds to the protein. The ribityl phosphate is exposed to bulk solvent but does make a salt bridge with the N-terminal amino group. Five ordered water molecules make bridging hydrogen bonds between the flavin and the protein, including hydrogen bonds to the side chains of Thr2, Tyr6, Asn190, Arg217 and Arg222. There is also an intramolecular hydrogen bond between the O4 hydroxyl of the ribityl group and the O2 carbonyl group of the isoalloxazine nucleus. Trp4 and Tyr221 make favorable stacking interactions with the isoalloxazine. The fatty acid binding pocket is made up of strand β_6 and the α_6 β_7 α_7 unit and is rich in aliphatic and aromatic side chains. Again, the acyl group of the fatty acid adopts a curved conformation in the hydrophobic binding pocket.

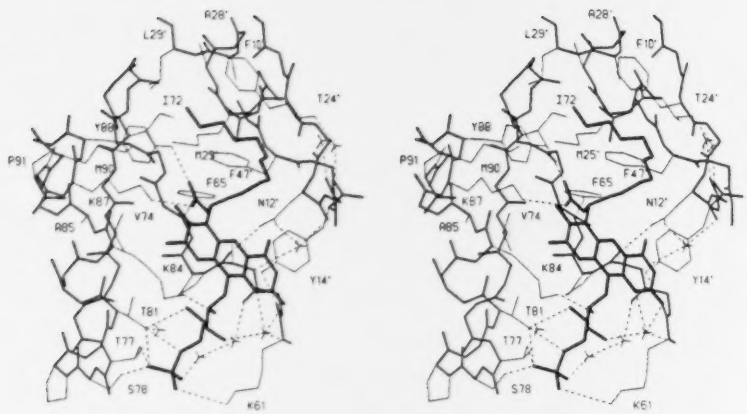


Figure 3.4a

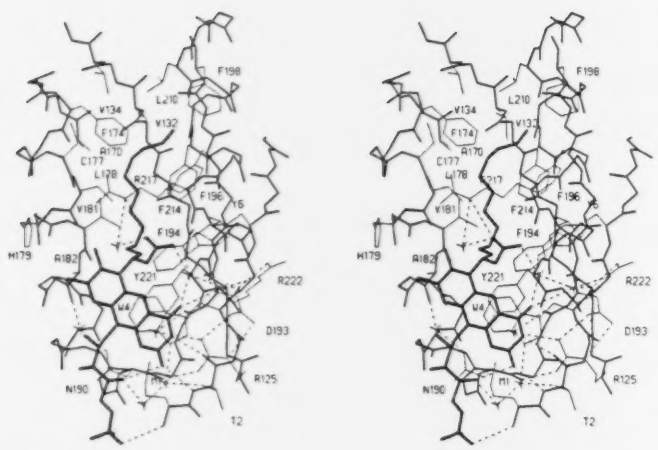


Figure 3.4b

chain of the myristic acid component is almost completely (96%) buried in a hydrophobic pocket at both sites. The pocket at the dimer interface is made up of hydrophobic side chains from the secondary structural units β_3 β_4 and β_1' α_1' β_2'. At the C-terminal site, hydrophobic side chains from secondary structural units β_6 α_6 and β_7 α_7 make up the fatty acid binding pocket. The fatty acyl group nestles between helices α_6 and α_7 and this likely explains why these helices are so obviously far apart in the structure. Strands β_6 and β_7 form the base of the binding pocket. As these cofactors can only be removed by methods that essentially denature NFP (O'Kane *et al.*, 1987; Raibekas, 1991), the hydrophobic interactions from the buried acyl side chains must make a major contribution to the binding free energies of the cofactors. The cofactor at the dimer interface appears more well ordered (B_{Ave} = 24.1 Å^2) than the cofactor at the C-terminal site (B_{Ave} = 44.8 Å^2) in our refined electron density maps (Figure 3.3). Much of the apparent thermal disorder for the second flavin is confined to the ribityl chain and the phosphate group. Only 40% of the solvent accessible surface area is buried for the ribityl phosphate at the C-terminus, versus 70% buried at the dimer interface site. There are nine hydrogen bonds to the ribityl phosphate for the flavin at the interface versus three for the C-terminal flavin (Figure 3.4). NFP is markedly nonfluorescent in solution and this fact is readily explained by the proximity of fluorescence quenching amino acids at both cofactor sites. Tyr14' stacks against the isoalloxazine ring system at the dimer interface (Figure 3.4a), and both Trp4 and Tyr221 stack against the isoalloxazine at the C-terminal site (Figure 3.4b).

DISCUSSION

Homology with Luciferase

From Figure 3.1b, it is evident that several of the conserved residues in the luciferase subunits and NFP cluster in two regions of the NFP molecule. One is located where helices α_2 and α_3 pack against strands β_5 and β_6. Here, the hydroxyl group of Tyr110 (on α_3) forms a hydrogen bond to the carbonyl oxygen of Met90 (on the loop connecting α_2 and β_5). Also in this region, the $N^{\delta 1}$ imidazole nitrogen of conserved His126 forms a hydrogen bond to the amide nitrogen of Leu92 (on β_5). The side chains of conserved Ala86 (on α_2) and His126 (on β_6) make favorable van der Waals contacts with the side chain of Tyr110. The Ala76 methyl group makes van der Waals contacts with the side chains of Leu92 and Phe94 (strand β_5), which align with bulky aliphatic residues in all luciferase sequences (Figure 3.1a). This cluster of conserved residues

could anchor an equivalent helix in luciferase to the NFP helix α_2 in a suitable position to interact with the ribityl phosphate group at the luciferase dimer interface. The hydroxyl group of Ser78 receives a helix-capping hydrogen bond from the Thr81 NH group (Figure 3.4b), and it also contributes a hydrogen bond to the ribityl phosphate group. This serine is the only residue conserved in both luciferase subunits and NFP that actually interacts with the NFP cofactor.

There is also a clustering of conserved aromatic residues at the C-terminal opening of the β-barrel. Conserved Trp96 (at the loop connecting β_5 and α_3) makes favorable van der Waals interactions with the side chain of residue 152 (on α_4) which is always a tyrosine or phenylalanine. Trp96 also interacts with His50 (on the β-meander connecting β_2 and β_3); this is a histidine in all but two of fifteen available luciferase and NFP sequences. Residue 51, which is histidine in all published sequences except *P. phosphoreum* NFP where it is a tyrosine, makes van der Waals contacts with Phe13 and Tyr152, both conserved aromatic groups.

Given the high degree of homology between the luciferase β subunit and NFP (Figure 3.1a), the common regions of these molecules likely share a similar three-dimensional fold. The strong homology observed between luciferase α and β subunits also suggests both molecules contain a fold similar to that observed in NFP. Supporting this hypothesis is the observation that residues strongly conserved in all three molecules cluster in two distinct regions of NFP (Figure 3.1b). Interestingly, one of these conserved clusters corresponds to a pocket formed by the loops connecting the carboxy ends of β-strands to their adjoining α-helices. This is equivalent to the structural arrangement of the active site regions in all α/β barrel enzymes (Branden & Tooze, 1991). This observation compels us to suggest that the luciferase active site has a similar location to active sites found in α/β barrel enzymes.

Supporting this hypothesis are site specific mutagenesis experiments that have shown His44 and His45 in the *Vibrio harveyi* luciferase α subunit are critical for full enzymatic activity (Xin *et al.*, 1991). These are the two histidines that align with His50 and His51 in *P. leiognathi* NFP, and lie in a loop at the carboxy-terminus of strand β_2 (Figure 3.1b). This cluster of conserved residues could be in close proximity to, or indeed, could be part of the luciferase active site. A twenty-seven amino acid α subunit insertion (Figure 3.1a), thought to be near the luciferase active site (Baldwin & Ziegler, 1992) would also map to the C-terminus of the barrel-forming β-strands, between

helices α_4 and α_5. Both luciferase and NFP are dimeric proteins, and from our alignment it appears that a significant portion of this interface (residues 8 to 72 in NFP) is conserved in the luciferase subunits (Figure 3.1a). However, probably only the major secondary structural elements in this region (helices α_1, α_2 and strands β_1, β_3) are conserved in luciferase and not the dimer interface. It is difficult to envision a similar interface in luciferase due to the presence of a 100 amino acid insertion in both luciferase subunits, somewhere between amino acids 60 and 70 in the NFP structure. Genetic evidence suggests that the luciferase β subunit and NFP arose from the luciferase α subunit by two consecutive gene duplication events (O'Kane & Prasher, 1992). The structure of NFP presented here gives valuable information pertaining to the structure of the luciferase subunits, and rationalizes the conservation of many amino acid residues in luciferase. Coupled with mutagenesis experiments and molecular modeling, these results should shed new light on luciferase structure and function.

MATERIALS AND METHODS

Structure Determination

Crystals of purified NFP from *P. leiognathi* strain S1 were grown as previously described (Chapter 2; Moore *et al.*, 1992). The crystals belong to space group $C222_1$ with one NFP monomer per asymmetric unit. Tetragonal crystals of the NFP from *P. phosphoreum* have been reported (Kita *et al.*, 1991). All X-ray intensity data were collected on a twin San Diego multiwire detector system (Xuong *et al.*, 1985) at the University of Alberta; no special procedures were used to enhance the collection of Bijvoet pairs. Data were processed using the San Diego software (Howard *et al.*, 1985). Heavy atom positions were determined directly from the difference Patterson maps and were refined using MLPHARE (Otwinowski, 1990). The only derivatives found were ethyl mercury phosphate and methyl mercury chloride. Each heavy atom bound at the same single-occupancy site to the sulphydryl group of Cys177 on helix α_6 with similar thermal parameters and near unit occupancy. Hence each derivative produced nearly identical phasing information (see Table 3.1). However, we chose to use both derivative data sets in the heavy atom refinement and phasing as maps calculated using phases derived from both data sets were somewhat more interpretable than maps computed from either data set alone. The ethyl mercury phosphate data was nearly complete, had a high phasing power and likely dominated the phasing. The available anomalous scattering signal was used in the refinement as the top three peaks in the anomalous Patterson

synthesis corresponded to the expected three heavy-atom Harker vectors. Maps were initially computed to 2.8 Å resolution for both heavy atom enantiomorphs and solvent flattened (Wang, 1985). The electron density map associated with one of the enantiomorphs contained stretches of right-handed α-helices. From this same map, it was obvious that the NFP dimer was located on the 0,y,1/4 twofold axis. A discontinuous model of 130 residues of polyalanine was built into the solvent flattened electron-density map using FRODO (Jones, 1978) and BONES (Jones & Thirup, 1986). Partial structure factors and phases were calculated for this model and combined to 2.6 Å resolution with the initial MIR phases using the program SIGMAA (Read, 1986). The N-terminal 130 amino acids of the published *P. leiognathi* NFP sequence (Illarionov *et al.*, 1990) were built into the resulting electron density map (<FOM> = 0.62). Fifty additional residues of polyalanine were also added to the model. This second partial model was subjected to conjugate gradient energy minimization within XPLOR (Brünger *et al.*, 1987) and the crystallographic R factor of the resultant model was 40%. The phases computed from this partial structure model were again combined with the MIR phases to generate a much improved map (<FOM> =0.70) calculated to 2.4 Å resolution. The rest of the amino acid sequence was built into this map and the crystallographic R-factor dropped to 25% (all data to 2.4 Å, $|F_o| \geq 3\ \sigma\ |F_o|$ after one round of simulated annealing using the slowcool protocol in XPLOR (Brünger *et al.*, 1990).

Refinement

At this point, the flavin myristate cofactors were built into the electron density, and the structure was refined with alternating rounds of model building and PROLSQ (Hendrickson & Konnert, 1980). The current R factor is 17.8% for all data with $|F| \geq 2\sigma|Fo|$ between 10 and 2.3 Å resolution (87.6% complete) for the 228 amino acid polypeptide model with 2 flavin myristate cofactors and 102 water molecules. The overall geometry of the model is good: r.m.s. deviations in bond distances =0.013 Å, r.m.s. deviations in angle-related distances = 0.031 Å, r.m.s. deviations in planar 1-4 distances =0.019Å and r.m.s. planar deviations = 0.007 Å.

REFERENCES

Argos, P., *Prot. Eng.* **2**, 101-113 (1988).

Baldwin, T.O. and M.M. Ziegler, in *Chemistry and Biochemistry of Flavoenzymes*, Vol. III, F. Müller, Ed., pp. 467-530 (CRC, Boca Raton, 1992).

Baldwin, T.O., J.H. Devine, R.C. Heckel, J.-W. Lin and G.S. Shadel, *J. Biolumin. Chemilumin.* **4**, 326-341 (1989).

Branden, C. and J. Tooze, Introduction to Protein Structure, pp. 47-48 (Garland, New York, 1991).

Brünger, A.T., A. Krukowski and J.W. Erickson, *Acta Crystallogr.* **A46**, 585-593 (1990).

Brünger, A.T., J. Kuriyan and M. Karplus, *Science* **235**, 458-460 (1987).

Connolly, M.L., *J. Appl. Crystallogr.* **16**, 548-558 (1983).

Daubner, S.C., A.M. Astorga, G.B. Leisman and T.O. Baldwin, *Proc. Natl. Acad. Sci. USA* **84**, 8912-8916 (1987).

Hastings, J.W., C.J. Potrikus, S.C. Gupta, M. Kurfürst and J.C. Makemson, in *Advances in Microbial Physiology*, Vol. 26, A.H. Rose and D.W. Tempest, Eds., pp. 235-291 (Academic, New York, 1985).

Hendrickson, W.A. and J.H. Konnert, in *Computing in Crystallography*, R. Diamond, S. Ramaseshan and K. Venkatesan, Eds., pp. 13.01-13.23 (Indian Institute of Sciences, Bangalore, 1980).

Howard, A.J., C. Nielsen and N.H. Xuong, in *Methods in Enzymology*, Vol. 114, Diffraction Methods for Biological Macromolecules, Part A , H.W. Wyckoff, C.H.W. Hirs and S.N. Timasheff, Eds., pp. 452-472 (Academic Press, New York, 1985).

Illarionov, B.A., V.M. Blinov, A.P. Donchenko, M.V. Protopopova, V.A. Karginov, N.P. Mertvetsov and J.I. Gitelson, *Gene* **86**, 89-94 (1990).

Jones, T.A. and S. Thirup, *EMBO J.* **5**, 819-822 (1986).

Jones, T.A., *J. Appl. Crystallogr.* **11**, 268-272 (1978).

Kabsch, W. and C. Sander, *Biopolymers* **22**, 2577 -2637(1983).

Karplus, P.A. and G.E. Schulz, *J. Mol. Biol.* **195**, 701-729 (1987).

Kasai, S., K. Matsui and T. Nakamura, in *Flavins and Flavoproteins* , D.E. Edmondson and D.B. McCormick, Eds., pp. 647-650 (Walter de Gruyter, Berlin, 1987).

Kasai, S., S. Fujii, R. Miura, S. Odani, T. Nakaya and K. Matsui, in *Flavins and Flavoproteins*, B. Curti, S. Ronchi and G. Zanetti, Eds., pp. 285-288 (Walter de Gruyter, Berlin, 1991).

Kita, A., N. Kasai, S. Kasai, T. Nakaya and K. Miki, *J. Biochem.* **110**, 748-750 (1991).

Kraulis, P.J., *J. Appl. Crystallogr.* **24**, 946-950 (1991).

Lee, J., I.B.C. Matheson, F. Müller, D.J. O'Kane, J. Vervoort and A.J.W.G. Visser, in *Chemistry and Biochemistry of Flavoenzymes*, Vol. II, F. Müller, Ed., pp. 109-151 (CRC, Boca Raton, 1991).

Lindqvist, Y. and C.-I. Brändén, *J. Biol. Chem.* **264**, 3624-3628 (1989).

Mayhew, S.G., C.D. Whitfield, S. Ghisla and M.S. Jörns, *Eur. J. Biochem.* **44**, 579-591 (1974).

Meighen, E.A., *Microbiol. Rev.* **55**, 123-142 (1991).

Moore, S.A., M.N.G. James, D.J. O'Kane and J. Lee, *J. Mol. Biol.* **224**, 523-526 (1992).

O'Kane, D.J. and D.C. Prasher, *Mol. Microbiol.* **6**, 443-449 (1992).

O'Kane, D.J., J. Vervoort, F. Müller and J. Lee, in *Flavins and Flavoproteins* , D.E. Edmondson and D.B. McCormick, Eds., pp. 641-645 (Walter de Gruyter, Berlin, 1987).

O'Kane, D.J., V.A. Karle and J. Lee, *Biochemistry* **24**, 1461-1467 (1985).

Otwinowski, Z., *Am. Crystallogr. Ass.* 1990. Annual Meeting, New Orleans, Abstr. C04 (1990).

Raibekas, A.A., *J. Biolumin. Chemilumin.* **6**, 169-176 (1991).

Read, R.J., *Acta Crystallogr.* **A42**, 140-149 (1986).

Singer, T.P. and W.S. McIntire, in *Methods in Enzymology*, Vol. 106, Post-translational Modifications, F. Wold and K. Moldave, Eds., pp. 369-378 (Academic, New York, 1984).

Soly, R.R., J.A. Mancini, S.R. Ferri, M. Boylan and E.A. Meighen, *Biochem. Biophys. Res. Commun.* **155**, 351-358 (1988).

Wang, B.C., in *Methods in Enzymology*, Vol. 115, Diffraction Methods for Biological Macromolecules, Part B, H.W. Wyckoff, C.H.W. Hirs and S.N. Timasheff, Eds., pp. 90-112 (Academic Press, New York, 1985).

Watt, W., A. Tulinsky, R.P. Swenson and K.D. Watenpaugh, *J. Mol. Biol.* **218**, 195-208 (1991).

Xin, X., L. Xi and S.-C. Tu, *Biochemistry* **30**, 11255-11262 (1991).

Xuong, N.H., C. Nielsen, R. Hamlin and D. Anderson, *J. Appl. Crystallogr.* **18**, 342-350 (1985).

CHAPTER 4

COMMON STRUCTURAL FEATURES OF THE *luxF* PROTEIN AND THE SUBUNITS OF BACTERIAL LUCIFERASE: EVIDENCE FOR A $(\beta\alpha)_8$ FOLD IN LUCIFERASE[1]

INTRODUCTION

Bacterial bioluminescence is a fascinating example of light emission from a living organism. All light-emitting bacteria have a lux operon which is responsible for the luminous phenotype (Meighen, 1993; 1991). Most of these bacteria are marine in origin and many are symbionts in fish. A terrestrial bioluminescent bacterium, *Xenorhabdus luminescens*, has a symbiotic relationship with a parasitic nematode and its host moth. How these often highly selective symbioses arose, and what advantages they convey to both symbionts are intriguing biological questions. Perhaps more interesting is how such luminous pathways were developed by bacteria, as several of the lux proteins seem to have been borrowed from other metabolic pathways (e.g. from riboflavin biosynthesis) (O'Kane and Prasher, 1992; Richter *et al.*, 1992; Lee *et al.*, 1992; O'Kane *et al.*, 1991).

The luminescence reaction (scheme 1) is catalyzed by bacterial luciferase, a heterodimeric flavin mono-oxygenase that uses myristic aldehyde, $FMNH_2$ and an oxygen molecule to produce light and the oxidized products: myristate, FMN and water (Baldwin and Ziegler, 1992). Bacterial luciferase guards its secrets closely; very little detailed chemistry is known about what takes place at the active site, with the exception of a 4a-peroxy-4a,5-dihydroflavin formed during the initial stages of the reaction (Lee *et al.*, 1991). Structurally, little is known about luciferase, although the α and β subunits can be easily aligned; they share roughly 30% pairwise amino acid identity. This level of sequence identity suggests that the subunits likely share a very similar overall fold, although some structural details will certainly be distinct for the two subunits. In fact the α subunit contains a 27 amino-acid insertion not present in β. About 10% of the amino acids in our alignment of presently available sequences of α and β subunits (Moore *et al.*, 1993) are completely conserved. Hence the subunits are clearly evolutionarily related and likely descended from a common ancestor (O'Kane and Prasher, 1992).

---

[1]A version of this chapter has been accepted for publication (Moore & James, 1994). *Protein Science*.

$$H^+ + FMNH^- + O_2 + RCHO \rightarrow FMN + H_2O + RCO_2H + h\nu \qquad \text{Scheme 1}$$

When α subunits are aligned 48% of the amino acids in the six sequences are identical. This compares to 31% of identical residues in a β subunit alignment. Hence, α subunit sequences are more strongly conserved, supporting the proposal that the active site of the enzyme is mainly on the α subunit. In general, genetic and enzymological evidence indicates a strong participation of the α subunit in catalysis (Baldwin and Ziegler, 1992).

We have recently determined the three dimensional structure of NFP (the protein product of the *luxF* gene) from *Photobacterium leiognathi* (Moore *et al.*, 1993), another component of the *lux* operon that sequence analysis clearly shows is related to the luciferase subunits. The *lux*F gene has only been found in the genus *Photobacterium*. Pairwise sequence alignments show the predicted protein sequence of the *lux*F gene to be 30% identical to the luciferase β subunit sequences and about 16% to α (Soly *et al.*, 1988; Moore *et al.*, 1993). Many of the strongly conserved residues in a luciferase subunit alignment are also strictly conserved in the two available NFP sequences.

Here we use the refined 1.6 Å NFP structure (Chapter 5) and amino acid sequence data to analyze conserved structural features in this family of proteins. A partial model of the luciferase β subunit from *P. leiognathi* has been built to illustrate the structural similarity. We have paid particular attention to structural motifs that are apparently conserved in all luciferase subunit protein sequences. A domain of approximately one hundred amino acids of both luciferase subunits is not found in NFP. Our work suggests that this region of the molecule likely completes a $(\beta\alpha)_8$ barrel structure. Only a modified seven-stranded barrel is found in NFP.

Like luciferase, NFP is a dimer but it is a homodimer. In the NFP structure, each monomer of 228 amino acids folds into a seven stranded β/α barrel. Six of the seven major β-strands are parallel and form a central elliptical barrel (Chapter 3; Moore *et al.*, 1993). Five of these β-strands are part of β/α units. However, the strands are arranged in such a way that there is a large gap between strands β_2 and β_3. This "side" of the barrel lacks surface helices and makes up the majority of the dimer interface in the NFP molecule. Several water molecules are found in place of the missing eighth strand that would be required to make a "closed" β-barrel topology typical of triose phosphate isomerase (TIM), glycolate oxidase and other enzymes (Farber and Petsko, 1990).

Hence the structure of NFP is analogous to the structure of a TIM barrel with a "bite" taken out of the side of it. In NFP, there are five βα units and two additional β-strands. In one of these βα units, a large surface excursion containing two α-helices and a short β-strand between strand β_6 and helix α_6 of NFP interrupts the topology. This structural feature also appears to be present in both luciferase subunits. Another surface excursion connects strands β_2 and β_3. It is part of the dimer interface in NFP and contains a short β-hairpin. There does not appear to be an equivalent feature in the polypeptide fold in luciferase.

MATERIALS AND METHODS

A database search looking for protein-folds similar to the NFP fold was carried out using an algorithm that aligns distance matrices between different proteins (Holm and Sander, 1993). Glycolate oxidase (Lindqvist and Brändén, 1989; Lindqvist, 1989) was found to have a fold that most closely resembled that of NFP (Liisa Holm and Chris Sander, private communication). Atomic coordinates for glycolate oxidase were obtained from the Brookhaven Protein Data Bank (Bernstein *et al.*, 1977). The initial alignment and superposition were optimized within O (Jones *et al.*, 1991) and the BIOMOL program package. The new PhD secondary structural prediction algorithm (Rost *et al.*, 1992) was used to look for the missing secondary structural units of a $(\beta\alpha)_8$ barrel in the 100 amino acid insertion of the luciferase subunits. The PhD algorithm works best when there is a family of related sequences available for multiple sequence alignments. Since several protein sequences are available for each luciferase subunit, we thought predictions of the α and β subunits, in the region of the 100 amino acid insertion, would be informative. PhD is available through an EMAIL facility at EMBL-Heidelberg and we ran separate secondary structure predictions for the α and β subunits of bacterial luciferase using all sequences available in the database. Predictions in regions of the α and β subunits having reasonable similarity to NFP would serve as useful controls.

Initial coordinates for the *P. leiognathi* luciferase β subunit model were generated from the *P. leiognathi* NFP structure using FRODO (Jones, 1978). Exchanged side chains making close van der Waals contacts were adjusted manually within FRODO. When adjusting side chain conformations, close attention was made to residue type and secondary structural preferences for that particular group (Dunbrack and Karplus, 1993). Two single-residue deletions were made, at Gln18 and Lys141 in the NFP structure.

The resulting main-chain geometry was adjusted and optimized using the REFI option with FRODO. An insertion of four residues between Gln157 and Ala158 in the NFP structure, was also made within FRODO. Lastly, the region of NFP between residues 52 and 70 was deleted from the model. After most of the bad stereochemical contacts were relieved manually, the model was energy-minimized for 800 cycles within XPLOR using the Powell minimizer (Brünger *et al.*, 1987). Coordinates were not restrained during the refinement. The amino acid bond length and angle parameters of Engh and Huber were used in the energy minimization (Engh and Huber, 1991). The model was again adjusted in regions of poor stereochemistry and close van der Waals contacts. Special attention was given to conserved regions in order to maintain the structural features observed in NFP. Three more rounds of 200 cycles each of refinement followed by manual rebuilding were carried out. The stereochemistry of the resulting partial model was checked after each round of refinement with PROCHECK (Laskowski *et al.*, 1993).

RESULTS

P. leiognathi luciferase β subunit model

Our partial model of the luciferase β subunit from *P. leiognathi* includes 201 amino acids and mimics the core of the overall structure found in *lux*F (Figure 4.1). The model does not include the 100 amino acid segment that has no counterpart in NFP. Also, part of the dimer interface in NFP, including strands β_{2a}, β_{2b} and β_3 was omitted as there is little homology with the luciferase β subunit sequences in this region. The predicted model displays good overall stereochemical characteristics. All residues are found within allowed regions of the Ramachandran plot. Good hydrogen-bonding geometry is observed in secondary structural units. An analysis of the final model using PROCHECK puts all stereochemical criteria within acceptable windows for a protein structure experimentally determined at 2.0 Å resolution. Generally hydrophobic residues face the interior of the protein and hydrophilic residues face solvent exposed regions.

Evidence for a $(\beta\alpha)_8$ structure

In any comparison of related protein sequences the regions of greatest sequence similarity generally contain conserved structural motifs. In the case of NFP where the three dimensional structure is known and a relatively straightforward alignment with both luciferase subunits is available (Meighen, 1991; Moore *et al.*, 1993), structurally

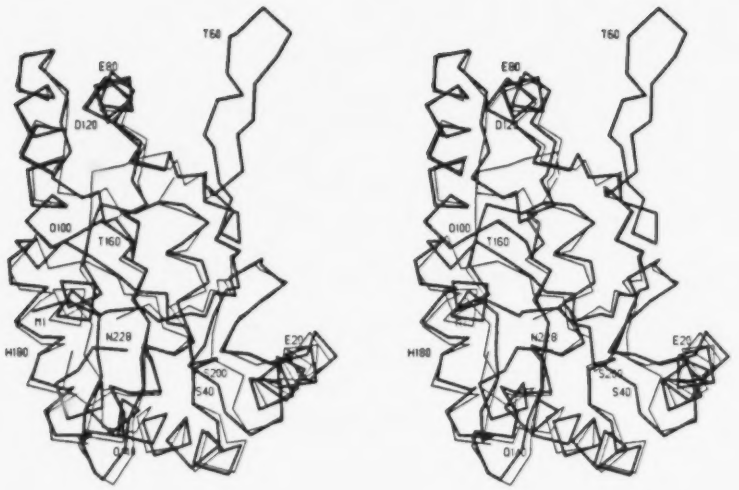


Figure 4.1. Superposition of the partial luciferase β subunit model onto the structure of NFP. Note that the C-terminal helix in the β subunit makes a rigid body rotation towards the penultimate helix. In NFP, a fatty acyl group is bound in a pocket between these two helices and splays them apart. This feature is not apparent in other TIM barrels and we have chosen to make the C-terminal helix of the β subunit in a conformation closer to the analogous helix in glycolate oxidase. The orientation is as in Figure 4.2. The NFP backbone is represented as a thick line with every 20th residue labeled. The β subunit model is represented as a solid thin line. The region corresponding to amino acids 52-70 of NFP is not included in the model.

conserved features in luciferase can be inferred. In fact, a pair of homologous proteins with 30% sequence identity would be expected to have a root mean square difference (rmsd) in main chain atom positions of ~1.5 Å (Chothia and Lesk, 1986). The actual rmsd between the main chain atoms of NFP and the β subunit model is 1.4 Å for 190 structurally equivalent amino acids.

In an exhaustive search of the Brookhaven protein data bank (Bernstein *et al.*, 1977) for folds similar to NFP, glycolate oxidase was found to have the structure that most closely matched that of NFP (Liisa Holm and Chris Sander, private communication). Glycolate oxidase is an FMN dependent enzyme that contains a $(\beta\alpha)_8$ barrel motif. This result was not surprising, as NFP looks more like an incomplete $(\beta\alpha)_8$ barrel than any other β/α folds.

Starting with the residue equivalences for glycolate oxidase as supplied by Liisa Holm and Chris Sander, we superposed NFP and glycolate oxidase. This crude superposition was then scrutinized for topologically equivalent regions and optimized. Our results are given in Table 4.1. Superposing the entire set of equivalent regions gave an rmsd of 2.4 Å for 347 main chain atoms. Superposing only C^α atoms gave an rmsd of 2.5 Å for 90 equivalent residues. In order to compare the central β-barrels of the two proteins, the helical regions were omitted from the superposition. The 130 equivalent main-chain atoms had an rmsd of 1.8 Å, and for a C^α superposition, the 29 equivalent atoms had an rmsd of 1.4 Å. These two superpositions are shown in Figure 4.1. The NFP barrel is more elliptical than that of glycolate oxidase, largely due to distortions of the barrel near the missing β-strand at the dimer interface in NFP. Strand β_2 of NFP especially, makes a poor superposition onto the second strand of the glycolate oxidase barrel.

Similarities to and differences from glycolate oxidase

The packing of aliphatic residues in the interior of the β-barrel of NFP is similar to the packing of side chains in glycolate oxidase and other $(\beta\alpha)_8$ barrel enzymes (Lesk *et al.*, 1989). In this family of proteins, hydrophobic residues pack in three layers with four residues in each layer. We had initially analysed the packing of hydrophobic amino-acid side chains in the central β-barrel of NFP without any prior knowledge of the structural alignment between glycolate oxidase and NFP. However structurally equivalent side chains from packing arguments are found to be equivalent to pairs of

Figure 4.2. Structural superposition of glycolate oxidase onto NFP.

(a) The equivalent C^{α}'s of NFP and glycolate oxidase are superposed to illustrate the similarity of the central β-barrel and some of the α-helices. The structurally equivalent regions for the superposition are given in Table 4.1. The rmsd in main chain atoms is 2.4 Å for 90 residues. NFP is drawn in thick lines and glycolate oxidase in thin lines. Every 20th residue of NFP is labeled. Breaks in the chains of NFP and glycolate oxidase are at amino acids 135 and 170 and 159 and 211, respectively. These regions of the structure do not form part of the $(\beta\alpha)_8$ barrel, and have been omitted for clarity.

Figure 4.2. (b) Drawn as Figure 4.2a. The main chain atoms for the central β-barrel portion of NFP and glycolate oxidase have been superposed. Only the β-sheet regions listed in Table 4.1 were used for the superposition. The rmsd in main chain atoms is 1.4 Å for 30 equivalent residues.

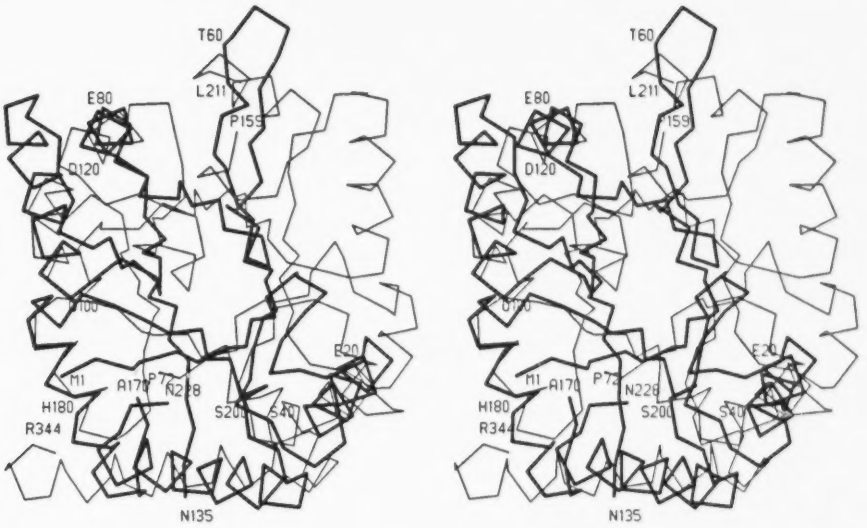


Figure 4.2a

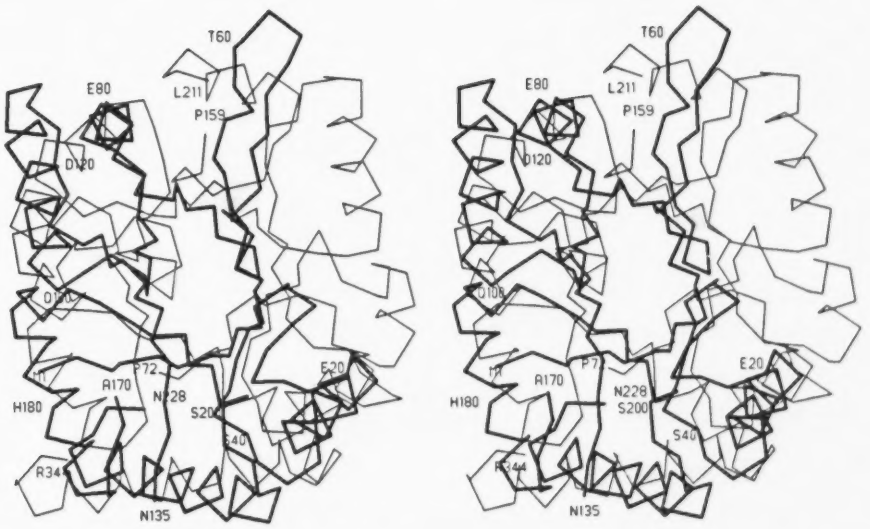


Figure 4.2b

Figure 4.2. (c) The relationship between the β-strands in a canonical $(\beta\alpha)_8$ barrel and the seven stranded NFP barrel is illustrated. The alignment of the barrel β-strands is based on an analysis of the packing of amino acids in the NFP β-barrel core (Table 4.2) and on the structural superposition of NFP with glycolate oxidase (Table 4.1).

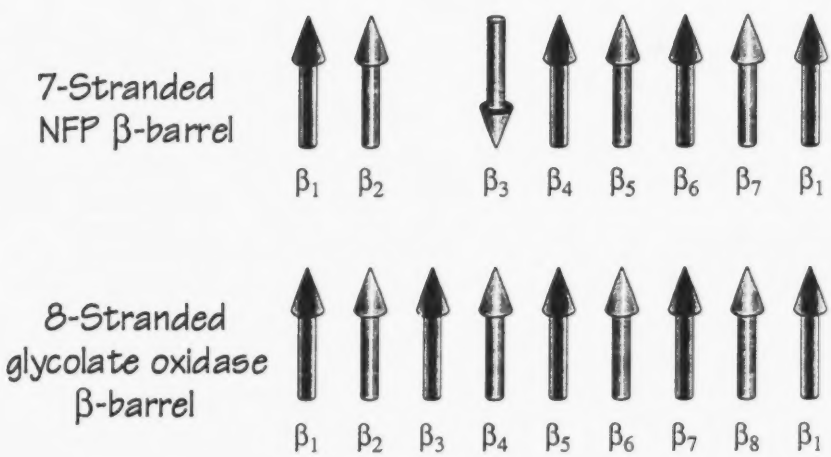


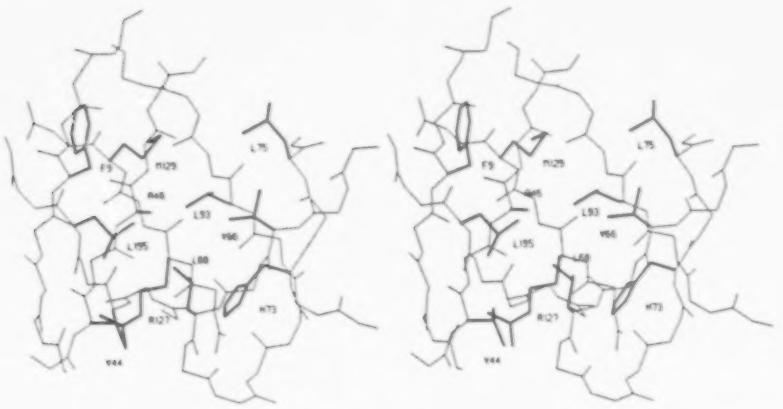
Table 4.1

Structurally equivalent polypeptide segments in NFP and glycolate oxidase

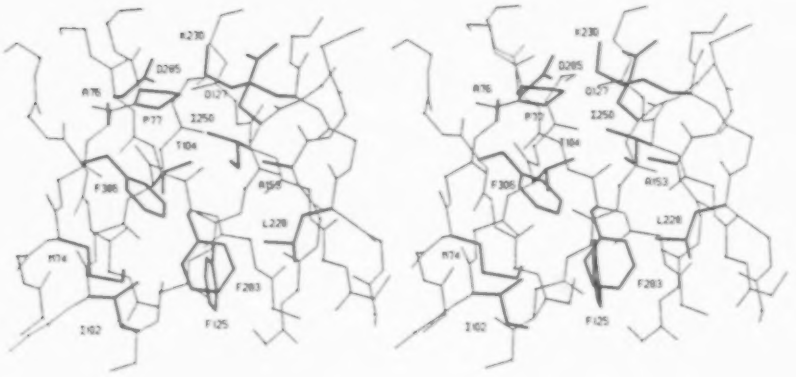
NFP	Glycolate Oxidase	Secondary Structure
N5-F10	P72-P77	β_1
N27-D36	G88-S97	α_1
D38-S40	A98-G100	α_1
D43-A46	T101-L104	β_2
Q71-L75	P226-K230	β_5
S78-K87	T234-Q243	α_5
M90-K95	A247-S252	β_6
S99-Q111	T265-Q277	α_6
V124-L130	I280-G286	β_7
C177-L188	R290-G301	α_7
D193-C197	G304-G308	β_8
F198-M201	V311-S314	
E205-S209	A321-K325	α_8

Structurally equivalent residues in NFP and glycolate oxidase are listed along with the secondary structural units of a canonical $(\beta\alpha)_8$ barrel with which they are equated. The aligned regions were taken from the superposition of the two molecules (Figure 4.2). The region Val311-Ser314 of glycolate oxidase makes a short helix between strand β_8 and helix α_8, and makes up a part of the FMN phosphate binding pocket in that enzyme. Interestingly, a similar segment is present in NFP.

Figure 4.3 Residues making up the β-barrel core in NFP and glycolate oxidase.



(a) A stereoview of the β-barrel core in NFP. Side chains are thick lines and the main chain β-strands are thin. The N-terminal opening of the barrel is at the bottom of the figure. The NFP barrel core residues are listed in Table 4.2.



(b) The side chain packing of the barrel core residues of glycolate oxidase is shown for comparison with NFP. The glycolate oxidase barrel core residues are listed in Table 4.2.

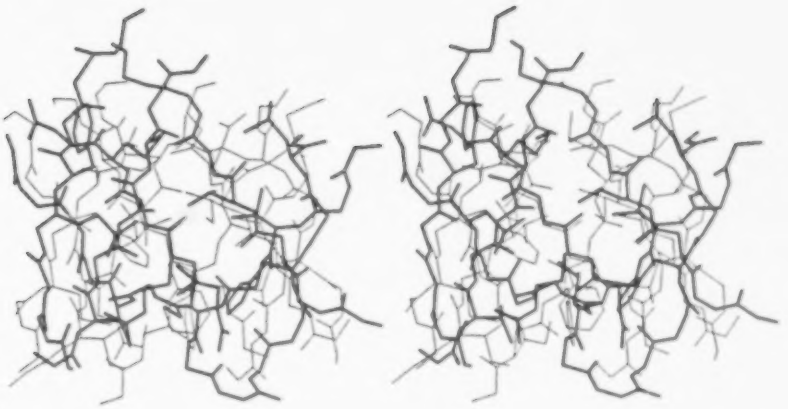


Figure 4.3. (c) The barrel core residues are overlaid for NFP (thick) and glycolate oxidase (thin) as per Table 4.2.

NFP

Glycolate oxidase

Figure 4.4. Stereo MOLSCRIPT (Kraulis, 1991) diagrams of the similar segments of NFP and glycolate oxidase looking from the direction of the NFP dimer interface. The glycolate oxidase coordinates are in an equivalent orientation to those of NFP. Regions in both molecules that don't contribute to the βα topology have been omitted for clarity.

Table 4.2. Amino acid residues making up the central core of the β-barrels in NFP and glycolate oxidase are compared. Residues making up the three layers of the β-barrel hydrophobic core in NFP obey the packing framework outlined by Lesk *et al.* (1989). Indeed when the structure of NFP was superposed onto that of glycolate oxidase (Table 4.1), there was a very good correspondence between the core forming residues of the two proteins. The matching core residues are given for each protein and the distances between the equivalent C^{α} atoms are also given. The distances are based on the superposition used for Figure 4.2b, where the rmsd for the β-barrel region is 1.4 Å (30 residues). The top, middle and bottom layers of the barrel core are delineated along with the parent β-strands of the glycolate oxidase $(\beta\alpha)_8$ barrel.

Table 4.2

Barrel core residues in NFP and glycolate oxidase

Strand Number (glycolate oxidase)	Glycolate Oxidase	NFP	Distance (Å)
Top Layer			
1	A76	F9	1.8
1	P77		
3*	Q127		
5	K230	L75	1.6
7	D285	M129	1.5
Middle Layer			
2	T104	A46	1.9
4	A153	V66	1.0
6	I250	L93	1.2
8	F2306	L195	0.6
Bottom Layer			
1	M74	G7	0.3
3*	F125	L68	3.5
5	L228	H73	0.7
7	F283	R127	1.6
2	I102	V44	1.3

* There is no strand equivalent to the third strand of a $(\beta\alpha)_8$ barrel in NFP.

amino acids in the NFP-glycolate oxidase structural superposition (see Table 4.2 and Figure 4.3). With NFP, strand β_3 of the central seven-stranded barrel is antiparallel and takes the place of strand 4 in glycolate oxidase. A large gap at the NFP dimer interface exists where strand 3 in glycolate oxidase is found. In NFP the middle layer of side chains is made up from even numbered strands if we number the NFP strands by taking account of the missing canonical third strand in glycolate oxidase (see Figure 4.3a). As a consequence of strand β_3 in NFP being antiparallel, it actually contributes a residue to the top and the middle layers, hence playing the role of the canonical strands 3 and 4 of a $(\beta\alpha)_8$ barrel (Lesk *et al.*, 1989). The side chain of Val66 on β_3 contributes to the middle layer of the barrel core (canonical strand 4) while the side chain of Leu68 on β_3 makes up part of the bottom layer (behaving as canonical strand 3). As odd numbered strands contribute to the top and bottom layers, and as there is no equivalent to a third strand in the NFP structure there are only three amino acid side chains in the top layer of the NFP barrel. In NFP the bottom layer, which is closer to the N-terminal opening of the barrel, is made up of residues Gly7, His73, and Arg127. Obviously Gly7 has no side chain, but the lack of a C^{β} atom leaves room for the side chain of Arg127. Both Val44 and Leu68 also contribute to the bottom layer in part making up for the vacancy left by Gly7 and the missing third canonical strand. The middle layer is made up of Ala46, Val66, Leu93 and Leu195. On the top layer at the C-terminal opening of the barrel are side chains from Phe9, Leu75, and Met129.

There are two structural groups in the family of $(\beta\alpha)_8$ proteins (Lesk *et al.*, 1989). In the first group, the top and bottom layers of the central hydrophobic core are made up of the side chains from odd-numbered strands of the β-barrel and the middle layer has side chains contributed from even-numbered strands. The other group has the middle layer composed of residues from odd-numbered strands of the β-barrel and the top and bottom layers are made up of side chains from the even-numbered strands. Both glycolate oxidase and NFP can be included in the former group of proteins.

Our superposition of NFP and glycolate oxidase suggests that NFP probably evolved from a progenitor containing a TIM barrel fold (Figure 4.2). To complete a $(\beta\alpha)_8$ fold, luciferase would need to have at least three additional α-helices and two β-strands relative to the NFP structure. This would require that the progenitor protein had an insertion relative to NFP somewhere following strand β_2 of NFP between amino acids 50 and 70. The first of these helices would likely follow strand β_2 of NFP and replace the β_{2a} β_{2b} hairpin at the NFP dimer interface. The first of the strands would be

found at the dimer interface of NFP. The second required α helix would again be found at the present NFP dimer interface. Following the second helix, the second required strand would replace strand β_3 of NFP and run parallel to the other strands of the barrel. The third required helix would again be in the region of the NFP dimer interface and it's C-terminus should meet up with the beginning of strand β_4 in NFP.

Is it possible for the luciferase subunits to adopt a fold similar to that observed in NFP and also contain a $(\beta\alpha)_8$ fold? We believe so. Secondary structure predictions for both luciferase subunits in regions of primary sequence homology to NFP were generally in good agreement with the experimentally determined structure of NFP (results not shown). This corresponds to two regions of NFP: namely the N-terminal stretch including β_1,α_1 and β_2, and the segment running from strand β_4 to the C-terminus of the molecule. Poor secondary structural agreement was observed between amino acids 56 and 72 of NFP and the corresponding segments of the luciferase subunits (Figure 4.5). This region is predicted to contain an α helix in both luciferase subunits, but contains a β-hairpin, and an antiparallel β-strand in NFP. This new helix which has no counterpart in NFP has been labelled α_{4L} in Figure 4.5. It also lies at the dimer interface in NFP. We feel that the structural similarity between NFP and the luciferase subunits will be poor in this portion of the NFP molecule. Directly following helix α_{4L} in the luciferase subunit amino acid sequences is an insertions of 100 amino acids relative to NFP. Figure 4.5 shows the secondary structure predictions in and around this 100 amino acid region for both luciferase subunits, as calculated by the PhD algorithm (Rost *et al.*, 1992).

The insertion must contain two β-strands and two α-helices to complete a $(\beta\alpha)_8$ fold. Analogous regions in other TIM barrel enzymes are generally 50-70 amino-acid residues in length. Hence the 100 amino acid region could easily accommodate the secondary structural features required to complete a canonical $(\beta\alpha)_8$ barrel. The insertion starts at residue 64 of the α subunit and ends at residue 167. The numbering is based on the *V. harveyi* α subunit amino acid sequence. Residues 72-78 are predicted to form a β-strand (β_{3L} in Figure 4.5). This strand would be equivalent to strand β_3 of glycolate oxidase, and has no counterpart in NFP. As in most other $(\beta\alpha)_8$ barrels, strand β_{3L} would likely run parallel to the rest of the β-strands in the luciferase subunits. Residues 83-97 of luciferase likely adopt a helical structure and are labeled as helix α_{3L} in Figure 4.5. The other β-strand (β_{4L}) that is required to complete a parallel eight-stranded barrel in luciferase, is predicted to occur at residues 101-105 in the luciferase sequence.

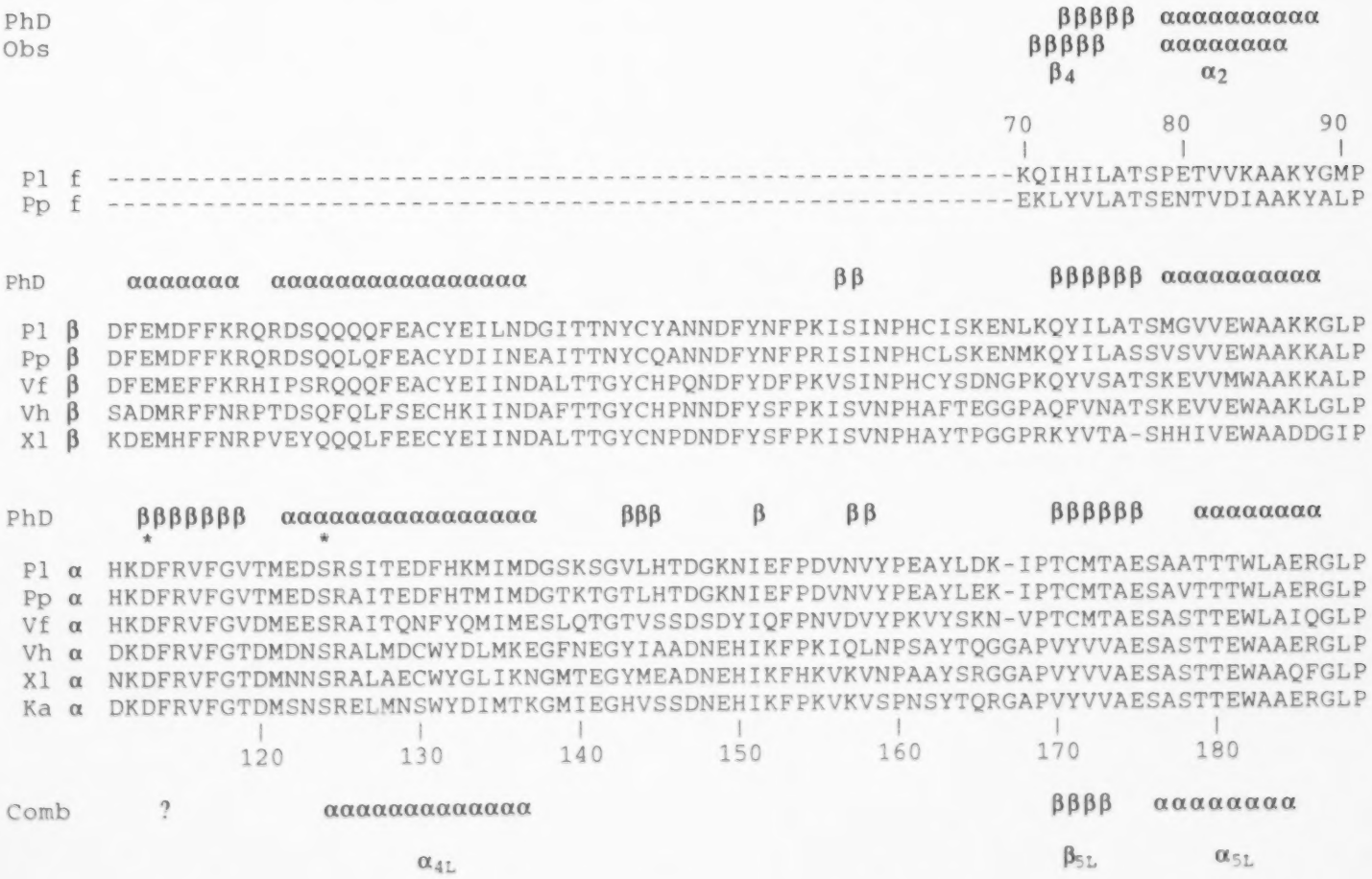
Figure 4.5. A secondary structure prediction for both luciferase subunits for the 100 amino acid region not found in NFP. Predictions are made for both groups of α and β subunit sequences. The α subunit amino acid sequence numbering is given at the bottom of the figure for the *V. harveyi* enzyme. The sequence numbering for *P. leiognathi* NFP is at the top of the figure. The secondary structure prediction and actual secondary structural elements for the determined structure of NFP (Chapter 3; Moore *et. al.*, 1993) are given at the top of the figure along with the NFP protein sequences bordering the region of the 100 amino acid insertion. Separate predictions for both the α and β subunit families are given above the respective sequences. The results for each subunit have been averaged to give a "mean prediction" at the bottom of the figure. The three predicted helical segments labeled α_{2L}, α_{3L} and α_{4L} likely correspond to the missing three helices of a $(\beta\alpha)_8$ barrel in the NFP structure. Strand β_{3L} likely fills up the gap at the dimer interface in NFP and strand β_{4L} likely replaces the antiparallel strand β_3 in NFP. A short stretch of 10 residues between positions 110 and 120 of the luciferase α subunit sequences appears to code for a short β-strand. This region is predicted to be an α-helix in the β subunit sequences. The region between amino acids 138 and 168 of both luciferase subunits display very little secondary structural preference but could form a β-hairpin or an extended loop region. This portion of the molecule would be at the N-terminal end of the β-barrel. Amino acid residues known to affect catalysis on the α subunit of the *V. harveyi* enzyme are marked with an asterisk. References for amino acid sequences used can be found in the legend to Figure 4.6.

Figure 4.5

Secondary structural prediction for the luciferase domain not found in NFP

PhD αα βββββ β ββ
Obs αααα ββββ βββ βββ ββββ
α1 β2 β2A β2B β3
40 50 60
Pl f IIDEDTSIYDVVAFSEHHIDKSYNDETKLAPFVSLG----------
Pp f IIDEDTSIYDVINIDDHYLVKKDSEDKKLASFITLG----------
PhD ββ βββββββββββ αααααααααα βββ βββββ ααααααααααααα ββββ
Pl β LVDKDKYHFKTAFVNEHHFSKNGIVGAPMTAASFLLGLTERLHIGSLNQVITTHHPVRIAEEASLLDQMSDGRFILGLSDCVS
Pp β LVDKDYKNFTTALVNEHHFSKNGIVGAPMTAASFLLGLTERLHIGSLNQVITTHHPVRIAEEASLLDQMSDSRFILGLSDCVN
Vf β LIDSTKYHFNTAFVNEHHFSKNGIVGAPI-AAGFLLGLTNKLHIGSLNQVITTHHPVRVAEEASLLDQMSEGRFILGFSDCES
Vh β YVDQ--LKFDTLAVYENHFSNNGVVGAPLTVAGFLLGMTKNAKVASLNHVITTHHPVRVAEEACLLDQMSEGRFAFGFSDCEK
Xl β YVDK--LNFEQILVYENHFSDNGVVGAPLTVSGFLLGLTEKIKIGSLNHIITTHHPVRIAEEACLLDQLSEGRFILGFSDCEK
PhD β βββββ ααα αααααααααα β βββββββ ααααααααααααα βββββββ
Pl α ASEE--LNFDTFWTLEHHFTEFGLTGNLYVACANILGRTKKLNVGTMGIVLPTAHPARQMEDLLLLDQMSKGRFNFGVVRGLY
Pp α ASEE--LNFDTYWTLEHHFTEFGLTGNLFVACANLLGRTTKLNVGTMGIVLPTAHPARQMEDLLLLDQMSKGRFNFGVVRGLY
Vf α ASEE--VGFDTYWTLEHHFTEFGLTGNLFVAAANLLGRTKTLNVGTMGVVIPTAHPVRQLEDVLLLDQMSKGRFNFGTVRGLY
Vh α ASEG--CGFDTVWLLEHHFTEFGLLGNPYVAAAHLLGATETLNVGTAAIVLPTAHPVRQAEDVNLLDQMSKGRFRFGICRGLY
Xl α ISEE--CGFDTVWLLEHHFTEFGLLGNPYVAAAYLLGATKKLNVGTAAIVLPTAHPVRQLEDVNLLDQMSKGRFRFGICRGLY
Ka α ASES--CGFDTAWLLEHHFTEFGLLGNPYVAAANLLGATKKLHVGTAAVVLPTAHPVRQLEDVNLLDQMSKGRFKFGICRGLY
40 50 60 70 80 90 100 110
Comb ββββ αααααααααα βββββββ αααααααααααααααα βββββ
β2 α2L β3L α3L β4L

Figure 4.5/contd.



This strand would be equivalent to strand 4 of the glycolate oxidase $(\beta\alpha)_8$ barrel and again would run parallel to the rest of the strands making up the luciferase β-barrel. Interestingly Cys106 of the *V. harveyi* α subunit lies at the C-terminus of this predicted strand and would be, by analogy with other $(\beta\alpha)_8$ enzymes, at the active site of luciferase. Cys106 of the *V. harveyi* α subunit is indeed known to affect the enzyme's function when alkylated, and it is also thought to form a thiohemiacetal with the substrate aldehyde under assay conditions. Residues 110-120 are predicted to adopt different secondary structures in the luciferase subunits and are likely not part of the $(\beta\alpha)_8$ topology. The last helix required to complete the $(\beta\alpha)_8$ barrel is predicted to lie between residues 124 and 136 of both luciferase subunits. This helix is labeled α_{4L} in Figure 4.5. A region of approximately 30 residues follows before the 100 amino acid insertion ends and the homology (and structural similarity) with NFP resumes. This 30 residue region is not predicted to have any significant secondary structural features, except for two short regions of β-structure. There are only a few strongly conserved amino acids in this 30 amino acid region (see Figure 4.5). Hence it likely forms a mostly extended structure that would lie at the N-terminal opening of the predicted $(\beta\alpha)_8$ barrel in luciferase.

One consequence of luciferase folding as a $(\beta\alpha)_8$ barrel, is that the dimer interface as observed in NFP cannot exist. Both luciferase and NFP are stable dimeric proteins; luciferase is an αβ heterodimer and NFP is a homodimer. Since the proteins are evolutionarily related, we initially thought both would have evolved containing essentially the same dimer interface. However this does not seem to be the case, as the insertion required to form a complete $(\beta\alpha)_8$ barrel in luciferase is structurally incompatible with the bulk of the NFP dimer interface. For whatever reason, when NFP diverged from the luciferase subunit genes, it gave up its progenitor's dimeric structure and found a totally new way of dimerizing. But this does point to the 100 amino acid insertion in the luciferase subunits as forming the dimerization interface. Firstly this region displays the highest sequence identity in a luciferase α/β subunit alignment. This suggests the interface could be two-fold symmetrical. It is also known that the β subunit of luciferase can form a homodimer in the absence of the α subunit (Friedland & Hastings, 1967) again arguing for a symmetrical dimeric structure.

Many $(\beta\alpha)_8$ barrel enzymes, including glycolate oxidase, utilize substrates that contain a phosphate group. Indeed both glycolate oxidase and luciferase bind FMN. All members of this subfamily contain a structurally conserved phosphate binding motif (Wilmanns *et al.*, 1991). Part of this motif is on a short α-helix, just before the eighth

helix of the $(\beta\alpha)_8$ barrel. In fact there is one turn of 3_{10} helix at the C-terminus of strand β_7 in NFP including residues 198 to 201 that overlaps reasonably well with residues 311 to 314 of the phosphate binding helix in glycolate oxidase. More interesting still is that the one invariant feature of these phosphate binding motifs is a glycine residue preceding the first amino acid of the helix. In NFP the equivalent residue is Cys197. Cys197 aligns with Gly326 of the *V. harveyi* luciferase α subunit. This amino acid is glycine in all other α subunit sequences, and mutation of Gly326 is known to affect the function of the enzyme (Johnston, 1993). A reasonable hypothesis is that the α subunit of luciferase binds FMN and that at least the mode of binding of the ribityl phosphate is similar to the binding of phosphate in glycolate oxidase and many other $(\beta\alpha)_8$ enzymes (Wilmanns *et al.*, 1991). If one assumes further that the overall orientation of FMN binding is similar in the luciferase α subunit and glycolate oxidase, it would put the pyrimidine moiety of the isoalloxazine ring in close proximity with two histidine residues (α44 and α45) that are critical for catalysis in *V. harveyi* (Xin *et al.*, 1991).

We believe both luciferase subunits belong to the large family of $(\beta\alpha)_8$ barrel enzymes which includes glycolate oxidase and triose phosphate isomerase. Residues on the luciferase α subunit in the 100 amino acid insertion that are known to affect catalysis are predicted to be found at the C-terminal region of β-strands. These residues include Cys106 and Asp113 in *V. harveyi*. Two histidine residues on the luciferase α subunit, also critical for light output (Xin *et al.*, 1991), map to the C-terminaus of strand β_2 in NFP (Moore *et al.*, 1993). Since all $(\beta\alpha)_8$ barrel enzymes have active sites constructed from residues in loops at the C-terminal ends of the β-strands, this information is entirely consistent with luciferase adopting such a structure.

Conservation of NFP helices in luciferase

α-Helices are often capped at the N-terminus by residues whose amino-acid side chains accept a hydrogen bond from the i+2 or i+3 main chain amide nitrogen atoms in the first turn of the helix (Baker and Hubbard, 1984; Richardson and Richardson, 1988). Helix capping residues include serine, threonine, aspartic acid, and asparagine. A perusal of the alignment of NFP with the luciferase subunits (Moore *et al.*, 1993) suggests that many of these N-terminal helix-capping residues are conserved. In addition, residues that terminate helices also are moderately well conserved in the luciferase subunits. These results are summarized in Table 4.3.

Table 4.3. Helix start and end points along with capping residues are listed for the seven α-helices in NFP. For each NFP helix, the main chain torsion angles are given for the listed residues. Amino acid residues found at these positions for the luciferase subunits are also tabulated. The sequence alignment used is the same for Figures 4.5, 4.6, 4.7, and 4.8. All numbers correspond to the NFP amino-acid sequence. Since there are eleven sequences for the luciferase subunits, more than one amino acid is found for a particular position in NFP.

Table 4.3

Helix start and end points in NFP and luciferase

	NFP			Luciferase		NFP			Luciferase	
NFP Helices	N-CAP in NFP	ϕ	ψ	Residues in β	Residues in α	C-CAP in NFP	ϕ	ψ	Residues in β	Residues in α
Helix α_1	E20	-77	151	T,S,V	T,S,D	D38	-100	-12		
Helix α_2	S78	-172	162	S	S	G89	66	25	G,A	G
Helix α_3	S99	-67	159	T,S,N	T,N,G	N118	59	50	N,G,D	G
Helix α_4	D137	-99	103	D	D,N	Q157	52	47	Y,H	N
Helix α_5						N169	-121	-158	H,N	N
Helix α_6	N173	-81	167	T,S,N	T	N189	57	24	G	G
Helix α_7	D203	-82	112	N,D	S,T					
	N-CAP	ϕ	ψ	Residues in β	Residues in α	C-CAP	ϕ	ψ	Residues in β	Residues in α

Helix α_1 in *P. leiognathi* NFP begins with a proline residue at position 21 (throughout this section, the *P. leiognathi* NFP protein sequence numbering will be used). A glutamate is found at the N-cap position and adopts the correct main-chain conformation to accept a hydrogen bond from the i+3 amide nitrogen (ϕ = -77, ψ = 151). However, in the NFP structure the glutamate side chain makes a salt bridge with His179 from a symmetry related molecule, and does not make the capping hydrogen bond. In both luciferase subunit sequences, there is a strong preference for a good N-cap residue at position 20 (Ser, Thr, Asp). Position 21 in the luciferase subunits is always a polar residue but it is not well conserved. The last residue in the helix, Asp36, is always aspartate in the β subunits, and always glutamate in the α subunits. The structural rationale for this conserved carboxylate is not evident. A plausible explanation though, is that a negatively charged Asp or Glu could interact unfavorably with the helix dipole, thereby destabilizing the helix. The C-cap at Asp38 has no counterpart in the luciferase sequences as there is a gap at this position of the alignment.

In helix α_2 of NFP, although Pro79 is the first residue in a helical conformation, it is not conserved in the luciferase sequences, and Ala is the only residue found in that position for the α subunit. Ser78 makes a good helix capping hydrogen bond in the NFP structure and lies in the far upper left hand region of the Ramachandran map (ϕ = -172, ψ = 162). A serine residue is strictly conserved in all luciferase sequences. At the C-terminus of the helix, Tyr88 lies just outside the accepted helical region (ϕ = -106, ψ = 9.5) and does not make an α helical hydrogen bond. In fact, the amide nitrogen atoms of both Tyr88 and Gly89 make reasonable i-3 hydrogen bonds. Gly89 is in a left-handed conformation and is a glycine residue in 10/13 of the *lux*A, B, F sequences. Tyr88 is less well conserved, but it is always a large bulky group in the luciferase α and β subunits. Most of the residues that precede the C-cap in NFP helices lie just outside the window defining α-helical residues. The majority of C-cap residues for the NFP helices are in a left-handed conformation (see Table4.3) and their amide groups (position i)generally make hydrogen bonds to the carbonyl group at position i-3. The majority of these C-cap residues are Gly, Asn or Asp, residues which are often found in a left-handed helical conformation.

Ser99 is the N-cap residue in helix α_3; this feature appears to be conserved in the luciferase sequences as only Ser, Thr or Asn are found at this position. Like many other helix-capping amino acids, Ser99 is in a β-conformation (ϕ = -67, ψ = 159). Another conserved residue near the N-terminus of helix α_3 is Arg103 (always Arg or Lys). The

main-chain carbonyl groups of the turn preceding helix α_3 are arranged so as to accept hydrogen bonds from the arginine guanidinium group. Residues in this turn are also strongly conserved. Hence the side chain of Arg103 appears not only to stabilize the start of helix α_3 in NFP, but also to anchor it in position. At the C-terminus of helix α_3, there is a strong preference for histidine or tyrosine at position 117 in luciferase. In NFP the (χ_1, χ_2) values are (-51, -45) for the phenylalanine residue at this position. This is similar to the side chain conformation of Tyr88 at the C-terminus of helix 2 (χ_1 = -52, χ_2 = -79). Like Tyr88, Phe117 is just outside the helical range of torsion angles (ϕ = -86.3, and ψ = -3.1) and has a χ_1 value near -60. In contrast, aromatic groups in helical regions usually have χ_1 values near 180°. It has been noted (Dunbrack and Karplus, 1993) that aromatic side chains just outside the helical range have a χ_1 value near -60°. Immediately following Phe117, residue 118 is in a left-handed helical conformation. It is Asp, Asn or Gly in all NFP and luciferase sequences.

In helix α_4, Asp137 makes a helix capping hydrogen bond, albeit an unusual one. Instead of making the typical i+3 hydrogen bond, it makes an i+2 hydrogen bond (actually a type III Asx turn) (Rees *et al.*, 1983) to the amide nitrogen of Thr139. Just previous to this Asx turn there is another strongly conserved type III Asx turn starting at Asn 135 (ϕ= -60, ψ=-27). Residue 135 is Asn or Asp in all luciferase sequences. Like most other capping residues, Asp137 is in a β-conformation (ϕ=-99, ψ=103), and in 12/13 luciferase subunit sequences this residue is Asp or Asn. Pro138, which is the first residue in a helical conformation (ϕ -53, ψ=-42), is not conserved in the luciferase subunits. The sequence of the loop preceding helix α_4 is reasonably well conserved especially in the β subunit of luciferase. Another apparently conserved structural feature in helix α_4 is a pronounced bend centred at position 150. Conserved aromatic residues at positions 148 and 152 anchor each helical segment relative to the rest of the protein by making hydrophobic interactions. Partially conserved Gly17 (7/13 sequences), at the C-terminus of strand β_1 in NFP, also makes close contacts with this helix bend region. The carbonyl oxygen of Ile147 and the C^{α} of Gly17 are only 3.3 Å apart. In addition, the amide nitrogen of Gly17 makes a favorable (3.1 Å) hydrogen bond with the carbonyl oxygen of Tyr148 in the NFP structure. The second segment of helix α_4 is broken by Gln157 which is in a left-handed conformation and makes an i-3 hydrogen bond to the carbonyl oxygen of Ser154. The helix end point is rather well conserved in the luciferase α subunit sequences, but not in the β subunits.

Helix α_5 is short and begins at Ile162, this position has an isoleucine in 8 of 13 luciferase and *lux*F sequences. However, the cap position at residue 161 is not conserved. Asn169 breaks the helix and is in a β-conformation (ϕ = -121, ψ = -158). It makes a good C-cap hydrogen bond from its amide nitrogen to the carbonyl oxygen of Ile166. The $N^{\delta 2}$ nitrogen of Asn169 makes an additional C-capping interaction by donating a hydrogen bond to the carbonyl oxygen of Ile165. Asn169 is strongly conserved as it is asparagine in 11/13 sequences. In the two exceptions, this residue is histidine and the imidazole side chain could easily make an analogous capping hydrogen bond with its $N^{\delta 1}$ atom.

Asn173 is the N-cap residue for helix α_6 and it is in a β-conformation (ϕ = -81, ψ = 167). The $O^{\delta 1}$ side chain atom of Asn173 makes an i+3 capping hydrogen bond. This residue is always asparagine, threonine, or serine but predominantly threonine (9/13 examples). In the luciferase α subunit sequences, the capping residue is always threonine and the predicted first helical residue at position 174 is invariantly proline. Proline often occupies the first position in α helices (Richardson & Richardson, 1988). Helix α_6 ends at Asn189, which is in a 3_{10} conformation. This residue is glycine in all luciferase sequences.

The last helix, helix α_7, is capped by Asp203 in NFP. The side chain carboxyl oxygen makes an i+3 hydrogen bond to the main-chain nitrogen of Asn206. Asp203 is in a β conformation (ϕ = -82, ψ = 112) and is a conserved capping residue (Asp, Asn, Ser, Thr) in all luciferase and NFP sequences. As there is variability in the C-termini of the luciferase subunit and NFP sequences, the end of helix α_7 could not be accurately predicted.

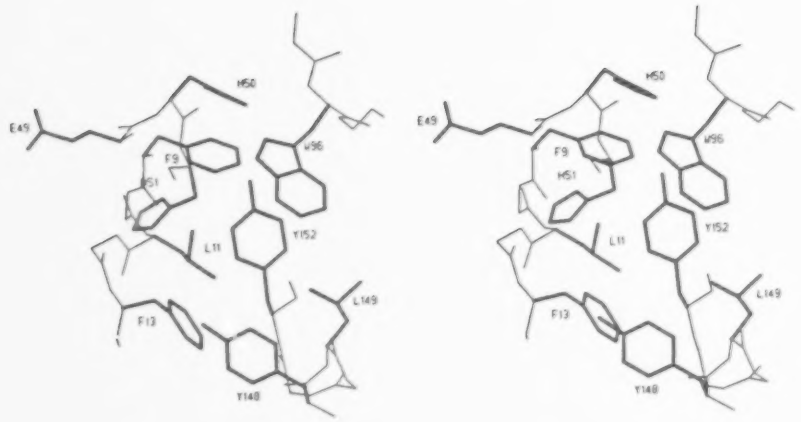
For each helix in NFP, amino acid sequence similarities argue for analogous helices in the luciferase subunits. In addition, the start and end points of the helices appear to be well-demarcated in the luciferase subunit amino acid sequences. These observations also support our hypothesis that there is extensive structural similarity between the luciferase subunits and NFP.

Conserved clusters involving surface loops

Another interesting feature of the NFP and luciferase subunit comparison is that clusters of conserved residues are near to, or are part of surface loops on NFP. This suggests a strong tendency to maintain key structural features in this family of proteins.

Figure 4.6. A stereo view of the cluster of conserved loops in NFP near Trp96. A section of the NFP-luciferase multiple sequence alignment corresponding to the conserved regions is also shown. Residues marked by a * are well conserved in both luciferase subunits and NFP. Residues marked with a $ are only conserved in NFP and the luciferase β subunit sequences. Conserved residues are also highlighted in bold. Sequences are: Pl F, luxF from *P. leiognathi*; Plα and Plβ, luciferase α and β subunits, respectively, from *P. leiognathi* (Illarionov *et al.*, 1990); PpF, luxF from *P. phosphoreum* (Soly *et al.*, 1988); Ppα and Ppβ, α and β subunit sequences from *P. phosphoreum* (Meighen, 1991); Vfα and Vfβ, the α and β subunit sequences from *V. fisheri* (Baldwin *et al.*, 1989); Vhα, the α subunit sequence from *V. harveyi* (Cohn *et al.*, 1985); Vhβ, the β subunit sequence from *V. harveyi* (Johnston *et al.*, 1986); Xlα and Xlβ, the α and β subunit sequences from *X. luminescens* (Szittner & Meighen, 1990; Johnston *et al.*, 1990); Ka α, the α subunit sequence for *Kryptophanaron alfredi* (Haygood, 1990). The amino acid sequence references also correspond to the sequences in Figures 4.5, 4.7 and 4.8.

Figure 4.6




```
                    10                  15                  50                  95                  100                 105
                    |                   |                   |                   |                   |                   |
                      *   $   *   $   *                       *   *   *           $   $   *       $                       *
P1  F   V F F L N F Y H V G Q   S E H H I   F K W D D S Q Q K R I E L
Pp  F   V F F V N F Y N K G Q   D D H Y L   F K W D D I N E E R L K L
P1  β   L F F L N F Q L K G -   N E H H F   Y R W S D T L A E K E N Y
Pp  β   L F F L N F Q P E N -   N E H H F   Y R W S D T L E D K E I L
Vf  β   L F F L N F Q K D G -   N E H H F   F K W E D N L E T K E R Y
Vh  β   L F F L N F M N S K -   Y E N H F   F R W D D S N A Q R K E Y
X1  β   L F F L N F I N S T -   Y E N H F   F K W D D S N D V R Y E Y

P1  α   N I C F S Y Q P P G -   L E H H F   L S W I I T T S E K K A Q
Pp  α   N I C F S Y Q P P G -   L E H H F   L S W I I T T S E K K A Q
Vf  α   N I C F S Y Q P P G -   L E H H F   L S W I I G T N E K K A Q
Vh  α   N F L L T Y Q P P E -   L E H H F   L S W I I N T H E K K A Q
X1  α   N F L L T Y Q P P Q -   L E H H F   L S W I I N T N E K K A Q
Ka  α   N F L L T Y Q P P Q -   L E H H F   L S W I I N N N E K K S Q
```

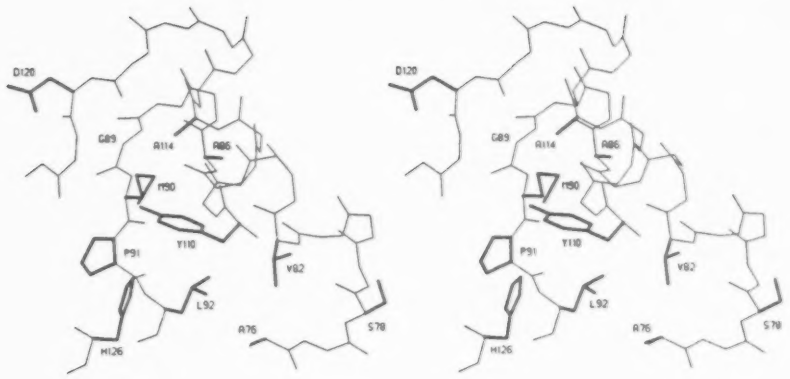
Conserved hydrophobic residues are not found in the barrel core, but between the barrel strands and the surface helices. There are three major regions in NFP where conserved residues in the luciferase subunits pack closely together. The residue numbering used below follows the *P. leiognathi* NFP amino acid sequence.

The first conserved cluster is located in the vicinity of the N-terminal β-strand region including residues Phe9 to Phe13, the surface loops Glu49 to Lys54 and Phe94 to Arg103, and residue Tyr152 on helix α_4 (Figure 4.6). Both conserved surface loops are at the C-terminal openings of the β-strands in NFP's central seven-stranded β-barrel. In addition, at least two residues in the loop corresponding to residues Glu49-Lys54 in the luciferase α subunit (His50 and His51) are crucial for bioluminescence light output (Xin *et al.*, 1991). Alignments for conserved amino acid segments in these regions for NFP and the luciferase subunits are also presented in Figure 4.6. Both Phe9 and Phe10 are conserved in the luciferase β subunits. Phe9 makes good van der Waals interactions with invariant Trp96 and with a hydrophobic residue at position 131. Leu11 is a moderately well conserved hydrophobic residue (Leu or Phe) and packs against Phe13, Trp96 and Tyr152. Phe13 is conserved in the β subunits and is a conserved tyrosine in the luciferase α subunits. Its side chain interacts with the side chains of conserved aromatic residues His51 and Tyr152. Glu49, His50 and His51 are in a 3_{10}-helical conformation on a surface loop and are well conserved in all luciferase sequences. The carbonyl oxygen of Glu49 makes a 3_{10} hydrogen bond to the amide nitrogen of Ile52.

A second conserved region includes helices α_2 and α_3 and strands β_5 and β_6 (see Figure 4.7). The critical feature here appears to be helix α_2, as it packs against the other structural units. It is found at the dimer interface in NFP. Loops preceding and following helix α_2 are well conserved, suggesting an important structural role for this portion of the molecule in NFP and also in the luciferase subunits. In NFP, residues on helix α_2 make several important interactions with one of the bound flavin cofactors. What function this part of the polypeptide would fulfill in the luciferase subunits is not clear. Conserved Ala76 makes a close contact, via its methyl group, with the side chain of Phe94. Residue 94 is always Phe or Tyr in the luciferase β subunits, and a Leu in the α subunits. Ser78 is invariant in all luciferase sequences and is at the N-cap position of helix α_2. In NFP Ser78 makes a strong hydrogen bond to the ribityl phosphate of one of the bound myristylated flavins (Chapter 3; Moore *et al.*, 1993). Thr81 is conserved in the luciferase α subunits and is a valine in the β subunits. This residue also forms a hydrogen bond to the flavin ribityl phosphate in NFP. Position 82 is always occupied

Figure 4.7. A stereo view of the conserved region including helices α_2 and α_3 and strands β_5 and β_6 of NFP. A multiple sequence alignment for the area shown is provided.

Figure 4.7



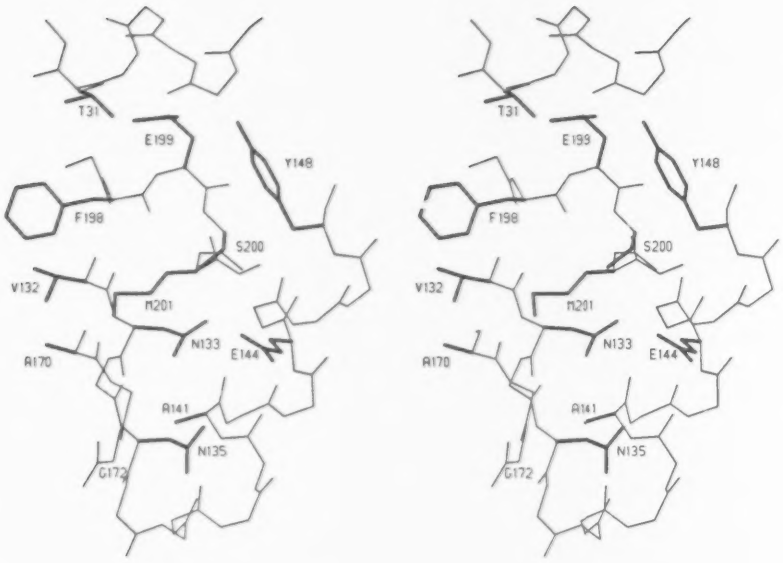

```
          75          80          85          90
          |           |           |           |
            *  $  *        *     $  *  $     *  *  *     $
Pl  F   L  A  T  S  P  E  T  V  V  K  A  A  K  Y  G  M  P  L  L  F
Pp  F   L  A  T  S  E  N  T  V  D  I  A  A  K  Y  A  L  P  L  V  F
Pl  β   L  A  T  S  M  G  V  V  E  W  A  A  K  K  G  L  P  L  T  Y
Pp  β   L  A  S  S  V  S  V  V  E  W  A  A  K  K  A  L  P  L  T  Y
Vf  β   S  A  T  S  K  E  V  V  M  W  A  A  K  K  A  L  P  L  T  F
Vh  β   N  A  T  S  K  E  V  V  E  W  A  A  K  L  G  L  P  L  V  F
Xl  β   T  A  T  S  H  H  I  V  E  W  A  A  D  D  G  I  P  L  I  F

Pl  α   T  A  E  S  A  A  T  T  T  W  L  A  E  R  G  L  P  M  V  L
Pp  α   T  A  E  S  A  V  T  T  T  W  L  A  E  R  G  L  P  M  V  L
Vf  α   T  A  E  S  A  S  T  T  E  W  L  A  I  Q  G  L  P  M  V  L
Vh  α   V  A  E  S  A  S  T  T  E  W  A  A  E  R  G  L  P  M  I  L
Xl  α   V  A  E  S  A  S  T  T  E  W  A  A  Q  F  G  L  P  M  I  L
Ka  α   V  A  E  S  A  S  T  T  E  W  A  A  E  R  G  L  P  I  I  L

          110         115         120                 125
          |           |           |                   |
            *           *           *       *       *   *
Pl  F   H  Y  Q  A  A  A  A  K  F  N  V  D  -  -  -  I  A  N  V  R  H  R  L
Pp  F   F  Y  N  A  S  A  S  K  Y  N  K  N  -  -  -  I  D  L  V  R  H  Q  L
Pl  β   R  Y  L  T  V  A  A  E  N  N  V  D  -  -  -  I  T  H  V  D  H  Q  F
Pp  β   R  Y  L  E  V  A  A  K  H  N  I  D  -  -  -  V  S  N  V  E  H  Q  F
Vf  β   L  Y  N  K  T  A  Q  Q  Y  G  V  D  -  -  -  I  S  D  V  D  H  Q  L
Vh  β   L  Y  H  E  V  A  Q  A  H  G  V  D  -  -  -  V  S  Q  V  R  H  K  L
Xl  β   R  Y  K  A  V  A  D  K  Y  D  V  D  -  -  -  L  S  E  I  D  H  Q  L

Pl  α   L  Y  N  E  I  A  A  E  H  G  H  D  -  -  -  I  H  N  I  D  H  S  M
Pp  α   L  Y  N  A  V  A  R  D  S  G  Y  S  E  E  Y  I  K  N  V  D  H  S  M
Vf  α   L  Y  N  E  I  A  T  E  Y  G  H  D  -  -  -  I  S  K  I  D  H  C  M
Vh  α   L  Y  N  E  V  A  Q  E  H  G  Y  D  -  -  -  V  T  K  I  D  H  C  L
Xl  α   L  Y  N  E  I  A  L  E  V  G  H  D  -  -  -  I  H  N  I  D  H  C  L
Ka  α   L  Y  N  E  I  A  L  E  H  G  H  D  -  -  -  V  S  N  I  D  H  C  M
```

by a β-branched residue in the luciferase sequence. Ala86 in NFP is invariant in both luciferase subunits, and its side chain packs against hydrophobic residues on the adjacent helix α_3, including the invariant Tyr110. Evidently the abundance of Val, Thr and Ala residues on this small helix allow it to pack closely to the neighboring helix α_3. Met90 and Pro91 are well conserved residues and form a loop leading from helix α_2 into strand β_5. The carbonyl oxygen of Met90 makes a strong (2.7Å) hydrogen bond to the hydroxyl group of invariant Tyr110 while the amide NH of Leu92, a conserved hydrophobe, makes a hydrogen bond to the unprotonated $N^{\delta 1}$ side chain nitrogen of conserved His126. In NFP, His126 makes close contacts with Tyr110 and Val124. Val124 is either Val or Ile in all luciferase sequences. Invariant Ala114 on helix α_3 is in close proximity with Ala86, and also makes good van der Waals interactions with the side chain of Ile121. Ile121 is either Ile or Val in all but one of the luciferase sequences and is the central residue of a conserved type I Asx turn (ϕ=-117, ψ=20). The previous residue, Asp120, makes the i+2 Asx hydrogen bond and is Asp in all but two luciferase and NFP protein sequences, where it is Asn or Ser. This constitutes a rich set of interactions involving helices α_2 and α_3 and also the loops that lead from these helices into the first residues of the ensuing β-strands, β_5 and β_6. These conserved interactions argue strongly for a similar αβαβ structural motif in the luciferase subunits.

A third region that exhibits remarkable conservation of sequence and therefore of tertiary structure among NFP and the luciferase subunits includes the C-terminus of strand β_6, helix α_4, and the loop between strand β_7 and helix α_7 (Figure 4.8). Also included is strand β_{6a} of NFP. Met201, which is conserved in all luciferase β subunit sequences makes hydrophobic interactions with Phe198. Phe 198 is conserved in all luciferase and NFP sequences and makes van der Waals interactions with Val132, a well conserved residue in the luciferase β subunit sequences. Phe198 and Met201 appear to act as a hydrophobic template, against which the hydrophobic face of the C-terminal α-helix packs. Both asparagines 133 and 135 on strand β_6 are strictly conserved in the β subunit sequences, and conservative amino acid substitutions are present in the α subunit sequences at these positions. As mentioned previously, Asn135 makes a conserved type III Asx turn. In NFP the side-chain carbonyl oxygen of Asn133 makes a hydrogen bond to the O^{γ} oxygen of Ser200 in a conserved loop between strand β_7 and helix α_7. Ser200 is invariant in the luciferase β subunit sequences. The $N^{\delta 1}$ nitrogen of Asn133 also makes a hydrogen bond to the $O^{\varepsilon 2}$ carboxylate oxygen of Glu144. Glu144 is located on the adjoining helix α_4 and is conserved in all luciferase β subunit sequences. Hence this

Figure 4.8. A stereo view of the conserved residues in NFP near Phe198. Amino acid residues in this region are much more strongly conserved in the luciferase β subunit sequences.

Figure 4.8




```
          25                  30                                  200
          |                   |                                   |
                              *                       * * $ $     $
Pl F   M S N A L E T L R        F C F E S M K D
Pp F   M N N A L E T L R        L C F E S M Q D
Pl β   L D N M I D T I A        L S F E S M K N
Pp β   L D N M I N T V S        L S F E S M K N
Vf β   L D N M V K T V T        L S F E S M A D
Vh β   I E E M L D T A H        M S F E S M E D
Xl β   I V R M Q E I T E        L S F E P M N D

Pl α   M D R F V R L G V        L G F E A N G S
Pp α   M D R F V R L G V        L G F E A N G S
Vf α   M D R F V R L G I        C G F E A N G T
Vh α   M K R L V N L G K        C G F E A N G S
Xl α   M K R L V K L G R        C G F E A N G T
Ka α   I K R L V N L G Q        C G F E A N G S

          130                 135                 140                 145                 150                 170
          |                   |                   |                   |                   |                   |
                $ $           *                   *                   $                   * *       *       * $     *
Pl F   L F V N V N D N P T Q A K A E L S I Y L - A D Y L S   S N A A G N F
Pp F   L H V N V N E A E T V A K E E L K L Y I - E N Y V A   S N V T G S Y
Pl β   L L V N I N P D R D I A K Q E M R D Y I - R G Y I A   Q H A V G T E
Pp β   L L V N L N H D R D V A H Q E A T A Y L - V S Y I A   Q H A I G T D
Vf β   V I A N L N S D R S T A Q E E V R E Y L - K D Y I T   E N A V G S H
Vh β   L L V N Q N V D G E A A R A E A R V Y L - E E F V R   E N A I G T Y
Xl β   I L V N Y N E D S N K A K Q E T R A F I - S D Y V L   E N A V G N Y

Pl α   F I C S V N E D P E K A E S V C R D F L I K L Y E S   L N P V G T P
Pp α   L I C S V D E D G K K A E D V C R E F L G N W Y D S   L N P V G T P
Vf α   Y I C S V D D D A Q K A Q D V C R E F L K N W Y D S   I N P V G T P
Vh α   Y I T S V D H D S N R A K D I C R N F L G H W Y D S   I N P V G T P
Xl α   Y I T S V D H D S I K A K E I C R K F L G H W Y D S   I N P V G T P
Ka α   Y I T S V D H N S N K A K D I C R D F L A H W Y D S   I N P V G T P
```

is likely a conserved hydrogen bonding interaction in the β subunits of bacterial luciferase. In addition, conserved Asn133, Asn135 and Glu144 form hydrogen bonds to a well-ordered water molecule at the turn between stand β_6 and helix α_4, in the NFP structure. This feature could be conserved in the luciferase β subunit. Ala141 is invariant in all luciferase subunit sequences; it makes a close contact with the carbonyl oxygen of Asn133. Invariant Gly172 is on strand β_{6a}, directly opposite to Ala141. On the loop between strand β_7 and helix α_7, the side chain carboxylate of invariant Glu199 makes a hydrogen bond to the O^{γ} oxygen atom of Thr31 (Figure 4.8). This interaction appears to be conserved in all β subunit sequences except for *Xenorhabdus luminescens* where the corresponding residue is Ile31. However Thr31 is on helix α_1 in NFP and the analogous i-4 residue is arginine in the *X. luminescens* β subunit sequence, and also in all α subunit sequences. This arginine could easily make a salt bridge with Glu199 of NFP.

CONCLUSIONS

We have presented structural and amino acid sequence information which suggest that both subunits of bacterial luciferase fold as $(\beta\alpha)_8$ barrels. The packing of the central β-barrel core of NFP, a homologue of the luciferase subunits, is found to be similar to the β-barrel core in glycolate oxidase. The structure of NFP is similar to glycolate oxidase in several respects, including the β-barrel and several α-helices. For 90 structurally equivalent residues, the rmsd in C^{α} atoms is 2.4 Å between NFP and glycolate oxidase. A region of 100 amino acids not found in NFP, but present in both luciferase subunits, is predicted to have enough secondary structural elements (two α helices and two β-strands) to complete the $(\beta\alpha)_8$ barrel proposed for the fold of the luciferase subunits. The third required helix comes before the insertion and replaces a region of β-structure in NFP. Amino acid sequences for the luciferase subunits are found to be generally consistent with the protein fold of NFP, especially at positions that demarcate the start and end points of α-helices. Several tertiary interactions involving loops at the C-terminal opening of the NFP β-barrel, also appear to be well conserved in luciferase. We suggest that the luciferase α subunit and glycolate oxidase bind the phosphate moiety of FMN in a similar manner at the N-terminus of a short α-helix between strand β_8 and helix α_8 of a $(\beta\alpha)_8$ barrel.

REFERENCES

Baldwin, T. O. & Ziegler, M. M. (1992). The biochemistry and molecular biology of bacterial bioluminescence. In *Chemistry and Biochemistry of Flavoenzymes* (Müller, F., ed.), vol. III, pp. 467-530, CRC, Boca Raton, FL.

Baker, E. N. & Hubbard, R. E. (1984). Hydrogen bonding in globular proteins. *Prog. Biophys. Molec. Biol.* **44**, 97-179.

Baldwin, T. O., Devine, J. H., Heckel, R. C., Lin, J. W. & Shadel, G. S. (1989). The complete nucleotide sequence of the *lux* regulon of *Vibrio fisheri* and the *lux* ABN region of *Photobacterium leiognathi* and the mechanism of control of bacterial bioluminescence. *J. Biol. Chemilumin.* **4**, 326-341.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer Jr., E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., Tasmui, M. (1977). The Protein Data Bank: A computer-based archival file for macromolecular structures. *J. Mol. Biol.* **112**, 535-542.

Brünger, A. T., Kuriyan, J. & Karplus, M. (1987). Crystallographic R factor refinement by molecular dynamics. *Science* **235**, 458-460.

Chothia, C. & Lesk, A. M. (1986). The relation between the divergence of sequence and structure in proteins. *EMBO J.* **5**, 823-826.

Cohn, D. H., Mileham, A. J., Simon, M. I., Nealson, K. H., Rausch, S. K., Bonam, D. & Baldwin, T. O. (1985). Nucleotide sequence of the *lux*A gene of *Vibrio harveyi* and the complete amino acid sequence of the α subunit of bacterial luciferase. *J. Biol. Chem.* **260**, 6139-6146.

Dunbrack, R. L. & Karplus, M. (1993). Backbone-dependent rotamer library for proteins. *J. Mol. Biol.* **230**, 543-574.

Engh, R. A. & Huber, R. (1991). Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Cryst.* **A47**, 392-400.

Farber, G. K. & Petsko, G. A. (1990). The evolution of α/β barrel enzymes. *TIBS* **15**, 228-234.

Friedland, J. & Hastings, J. W. (1967). Nonidentical subunits of bacterial luciferase: their isolation and recombination to form active enzyme. *Proc. Natl. Acad. Sci.* **58**, 2336-2342.

Haygood, M. G. (1990). Relationship of the luminous bacterial symbiont of the Caribbean flashlight fish, *Kryptophanaron alfredi* (family *Anomalopidae*) to other luminous bacteria based on bacterial luciferase (*lux*A) genes. *Arch. Microbiol.* **154**, 496-503.

Holm, L. & Sander, C. (1993). Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123-138.

Illarionov, B. A., Blinov, V. M., Donchenko, A. P., Protopopova, M. V., Karginov, V. A., Mertvetsov, N. P. & Gitelson, J. I. (1990). Isolation of bioluminescent functions from *Photobacterium leiognathi*: analysis of *lux*A, *lux*B, *lux*G and neighboring genes. *Gene* **86**, 89-94.

Johnston, T. C., Rucker, E. B., Cochrum, L., Hruska, K. S. & Vandegrift, V. (1990). The nucleotide sequence of the *lux*A and luxB genes of *Xenorhabdus* luminescens Hm and a comparison of the amino acid sequences of luciferases from four species of bioluminescent bacteria. *Biochem. Biophys. Res. Commun.* **170**, 407-415.

Johnston, T. C., Thompson, R. B. & Baldwin, T. O. (1986). Nucleotide sequence of the *lux*B gene of *Vibrio harveyi* and the complete amino acid sequence of the β subunit of bacterial luciferase. *J. Biol. Chem.* **261**, 4805-4811.

Johnston, T.C. (1993). Alignment of the α subunits of bacterial luciferase and sequence analysis of mutant luciferases showing altered kinetics. *Bioluminescence Symposium*, Maui, Hawaii. November 5-10, 1993. Abstr. Sp33.

Jones, T. A. (1978). A graphics model building and refinement system for macromolecules. *J. Appl. Cryst.* **11**, 268-272.

Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Cryst.* **A47**, 110-119.

Kraulis, P. J. (1991). MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.* **24**, 946-950.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283-291.

Lee, C. Y. & Meighen, E. A. (1992). The *lux* genes in *Photobacterium leiognathi* are closely linked with genes corresponding in sequence to riboflavin synthesis genes. *Biochem. Biophys. Res. Commun.* **186**, 690-697.

Lee, J., Matheson, I. B. C., Müller, F., O'Kane, D. J., Vervoort, J. & Visser, A. J. (1991). The mechanism of bacterial bioluminescence. In *Chemistry and Biochemistry of Flavoenzymes* (Müller, F., ed.), pp. 109-151, CRC, Boca Raton, FL.

Lesk, A.M., Brändén, C.-I. & Chothia, C. (1989). Structural principles of α/β barrel proteins: the packing of the interior of the sheet. *Proteins* **5**, 139-148.

Lindqvist, Y. & Brändén, C.-I. (1989). The active site of spinach glycolate oxidase. *J. Biol. Chem.* **264**, 3624-3628.

Lindqvist, Y. (1989). Refined structure of spinach glycolate oxidase at 2 Å resolution. *J. Mol. Biol.* **209**, 151-166.

Meighen, E. A. (1991). Molecular biology of bacterial bioluminescence. *Microbiol. Rev.* **55**, 123-142.

Meighen, E. A. (1993). Bacterial bioluminescence: organization, regulation, and application of the *lux* genes. *FASEB J.* **7**, 1016-1022.

Moore, S. A., James, M. N. G., O'Kane, D. J. & Lee, J. (1993). Crystal structure of a flavoprotein related to the subunits of bacterial luciferase. *EMBO J.* **12**, 1767-1774.

O'Kane, D. J. & Prasher, D. C. (1992). Evolutionary origins of bacterial bioluminescence. *Mol. Microbiol.* **6**, 443-449.

O'Kane, D. J., Woodward, B., Lee, J. & Prasher, D. C. (1991). Borrowed proteins in bacterial bioluminescence. *Proc. Natl. Acad. Sci.* **88**, 1100-1104.

Rees, D. C., Lewis, M. & Lipscomb, W. N. (1983). Refined crystal structure of carboxypeptidase A at 1.54 Å resolution. *J. Mol. Biol.* **168**, 367-387.

Richardson, J. S. & Richardson, D. C. (1988). Amino acid preferences for specific locations at the ends of α helices. *Science* **240**, 1648-1652.

Richter, G., Volk, R., Krieger, C., Lahm, H. W., Rothlisberger, U. & Bacher, A. (1992). Biosynthesis of riboflavin: cloning, sequencing and expression of the gene coding for 3, 4-dihydroxy-2-butanone 4-phosphate synthase of *Escherichia coli*. *J. Bacteriol.* **174**, 4050-4056.

Rost, B., Schneider, R. & Sander, C. (1993). Progress in protein structure prediction? *TIBS* **18**, 120-123.

Soly, R. R., Mancini, J. A., Ferri, S. R., Boylan, M. & Meighen, E. A. (1988). A new *lux* gene in bioluminescent bacteria codes for a protein homologous to the bacterial luciferase subunits. *Biochem. Biophys. Res. Commun.* **155**, 351-358.

Szittner, R. & Meighen, E. (1990). Nucleotide sequence, expression and properties of luciferase coded by *lux* genes from a terrestrial bacterium. *J. Biol. Chem.* **265**, 16581-16587.

Wilmanns, M., Hyde, C. C., Davies, D. R., Kirschner, K. & Jansonius, J. N. (1991). Structural conservation in parallel β/α-barrel enzymes that catalyse three sequential reactions in the pathway of tryptophan biosynthesis. *Biochemistry* **30**, 9161-9169.

Xin, X., Xi, L. & Tu, S.-C. (1991). Functional consequences of site-directed mutation of conserved histidyl residues of the bacterial luciferase α subunit. *Biochemistry* **30**, 11255-11262.

CHAPTER 5

1.6 Å RESOLUTION REFINEMENT OF THE NON-FLUORESCENT FLAVOPROTEIN FROM *Photobacterium leiognathi*

INTRODUCTION

Bioluminescent bacteria belonging to the genus *Photobacterium* characteristically produce large quantities of a non-fluorescent flavoprotein (NFP) under growth conditions favoring induction of the *lux* operon (O'Kane *et al.*, 1987; Kasai *et al.*, 1987). The *lux* operon contains genes required for bioluminescence including those coding for the α and β subunits of bacterial luciferase, a heterodimeric flavin mono-oxygenase that uses myristic aldehyde as substrate. The enzymes necessary for biosynthesis of myristic aldehyde are also encoded on the *lux* operon. These operons from bacteria belonging to the *Photobacterium* genus also have a gene, *lux*F, that codes for a 228 or 231 residue polypeptide depending upon the species (Soly *et al.*, 1988; Illarionov *et al.*, 1990). The predicted amino-acid sequence of this polypeptide is identical to the determined N-terminal amino acid sequence of *P. leiognathi* NFP (D.J. O'Kane, unpublished results; Raibekas, 1991). Since NFP is coded by the *lux*F gene on the *lux* operon, it is clear to see that this protein is co-induced with bacterial luciferase. More importantly, the sequences of both subunits of bacterial luciferase can be aligned with NFP (Soly *et al.*, 1988; O'Kane & Prasher, 1992; Moore *et al.*, 1993), suggesting these proteins evolved from a common ancestor.

NFP is not essential for bioluminescence (Meighen, 1991), but its possible effect on light emission in the host organism has not been studied. It also has no obvious function other than that it binds two molecules of a highly unusual covalent adduct of flavin mononucleotide (FMN) and myristate (O'Kane *et al.*, 1987; Kasai *et al.*, 1991; Moore *et al.*, 1993). Interestingly, both myristate and FMN are products of the bioluminescence reaction. This prompted us to hypothesize that the NFP chromophore is actually an inhibitory side-product of the luciferase-catalyzed reaction (Moore *et al.*, 1993). We have previously determined the three dimensional structure of NFP from *Photobacterium leiognathi* at 2.3 Å resolution. Crystals of the NFP homologue from *Photobacterium phosphoreum* have been reported (Kita *et al.*, 1991). Our structure determination showed that the C6 carbon atom of the flavin isoalloxazine ring makes a covalent bond with the C3' methine carbon atom of the fatty acid. The attachment

appears to be stereoselective as only the S isomer is seen at C3' of the fatty acid (Moore *et al.*, 1993)*. This suggests the involvement of an enzyme in the formation of the isoalloxazine-fatty acid bond. We now extend our structural analysis of NFP to 1.6 Å resolution using data collected on imaging plates with synchrotron radiation.

MATERIALS AND METHODS

Data collection and processing

Single crystals of NFP (space group $C222_1$) were grown as described previously (Moore *et al.*, 1992). Crystals were mounted in a mother liquor of 45% saturated ammonium sulphate, 100 mM sodium citrate, pH 5.5, in sealed glass capillaries prior to X-ray exposure. High resolution X-ray diffraction data were collected on Fuji imaging plates using a Weissenberg camera (Sakabe, 1983; Sakabe, 1991) at beamline BL-6A2 at the Photon Factory in Tsukuba Japan. Two crystals were used for the data collection and a total of 34 imaging plates were recorded. The space group and unit cell dimensions were identical to those published previously (Moore *et al.*, 1992). Details of the data collection have been summarized in Table 5.1.

Diffraction intensities were recovered from the imaging plate data using WEIS (Higashi, 1989; Fields *et al.*, 1992) operating on a Silicon Graphics workstation. Images from each of the two crystals were processed separately and subjected to a multiple image refinement. In multiple image refinement, all of the imaging plates corresponding to one crystal mount are refined simultaneously with the same missetting angles, camera geometry and unit cell dimensions. Integrated intensities from the two separate datasets were then merged and scaled together using a non-linear least squares procedure (Hamilton *et al.*, 1965). The final unit cell dimensions were averaged from the two datasets. Merging statistics for the combined data are given in Table 5.2. An absolute scale factor and an overall B-factor for the data were calculated using ORESTES (Thiessen & Levy, 1973). Approximately 81% of the theoretically observable reflections between infinity and 1.6 Å resolution were recorded. On average, each individual reflection was measured three times. The highest resolution shell (1.67-1.60 Å) was 66% complete. Completeness of the data as a function of resolution is given in Table 5.3.

---

*In our original 2.3 Å structure determination, we had determined the correct absolute configuration about C_3, but incorrectly named it as the R enantiomer.

Table 5.1

Photon factory imaging plate data for NFP

	Crystal #1	Crystal #2
Wavelength (Å)	1.00	1.00
Film Radius (mm)	286.5	286.5
Number of IP's	18	16
ϕ_x (°)	-1.77	-2.60
ϕ_y (°)	0.19	0.62
ϕ_z (°)	-1.09	-8.00
rms dR(mm)	0.31	0.40
rms dS(mm)	0.16	0.21
a (Å)	57.05	56.94
b (Å)	92.24	92.25
c (Å)	99.42	99.25
Rotation axis	a*	c*
X-ray axis	b*	a*

Table 5.2

Merging statistics for NFP data

a (Å)	57.00
b (Å)	92.24
c (Å)	99.33
Integrated intensities	84138
Unique measurements	28223
Resolution limit (Å)	1.6
R_{merge}‡	0.059
Scale factor	21.93
B_{iso} ($Å^2$)	16.07

‡ $$R_{merge} = \frac{\sum_{hkl} \sum_{i} |I_{i,hkl} - \langle I \rangle_{hkl}|}{\sum_{hkl} \sum_{i} I_{i,hkl}}$$

Table 5.3

Completeness of NFP diffraction data as a function of resolution

Resolution Shell (Å)	Possible Refls.	Integrated Refls	% Obs.
Infinity - 3.33	4031	3650	90.5
3.33 - 2.58	4473	4136	92.5
2.58 - 2.24	4492	3970	88.4
2.24 - 2.04	3989	3413	85.6
2.04 - 1.89	4334	3484	80.4
1.89 - 1.77	4627	3489	75.4
1.77 - 1.67	4936	3491	70.7
1.67 - 1.60	3954	2590	65.6
Totals	34,836	28,223	81.0

Least squares refinement

The starting model for restrained-parameter least-squares refinement was the 2.3 Å M.I.R. structure of NFP with all water molecules removed (Moore *et al.*, 1993; Chapter 3). A SIGMAA weighted electron density map was calculated (Read, 1986) and the model was fitted to the new map within FRODO (Jones, 1978). After initial rebuilding, the model was refined against all data with $|F_o| \geq \sigma|F_o|$ between 10.0 and 1.60 Å resolution in PROLSQ (Hendrickson & Konnert, 1980). Solvent sites were chosen as peaks being larger than three times the RMS value of an $m|F_o| - D|F_c|$ SIGMAA weighted map. Nearly all of the solvent sites chosen in the 2.3 Å structure of NFP came back as strong peaks in the new difference maps and refined as well ordered solvents (occupancy near 1.0 and $B < 30 Å^2$). The ideal geometries for the FMN and myristic acid moieties were defined as described previously (Moore *et al.*, 1993; Chapter 3). After ninety-three cycles of PROLSQ with model refitting and addition of solvent molecules after each round of 8-10 cycles, the resultant crystallographic R-factor was 0.175 for data between 10.0 and 1.60 Å resolution. Refinement statistics are summarized in Table 5.4.

All but two of the 228 amino acids of the NFP polypeptide are well defined in our final $2|F_o| - |F_c|$ electron density maps. Residues Glu159 and Thr160 have poorly associated electron density in the maps. A Ramachandran plot of the final model is shown in Figure 5.1. Ser161, is found in a disordered region and is the only residue in a disallowed conformation. Other portions of the model associated with breaks in the electron density are given in Table 5.5. Most of the breaks in the main-chain density are associated with C^{α}-C bonds. The C^{α}-C bond is the longest in the main chain of the polypeptide (1.54Å) and the bond length is very close to the resolution limit of this analysis (1.60Å). One could hypothesize that the observed electron density breaks are in fact the result of the atomicity of the electron distribution and at a resolution of 1.60Å, we are beginning to separate the overlap of the contributing C^{α} and carbonyl carbon atomic electron distributions.

We have modeled four amino-acid side chains with alternate conformations in our refinement. They are Ser55, Lys61, Ser67 and Ser209 and all make good hydrogen bonds in both conformations. A total of 191 solvent molecules and one sulphate group have been included in the model.

Table 5.4

Refinement statistics for P. leiognathi NFP

Resolution Range (Å)	10.0 - 1.60
Number of reflections	27451
Completeness of data	78.8%
R-factor	0.175
Total number of atoms (excluding solvent)	1948
Total number of solvent atoms	191
rms deviations from ideal stereochemistry:	
bond distances	0.018 Å
angle-related distances	0.032 Å
peptide planes	0.014 Å

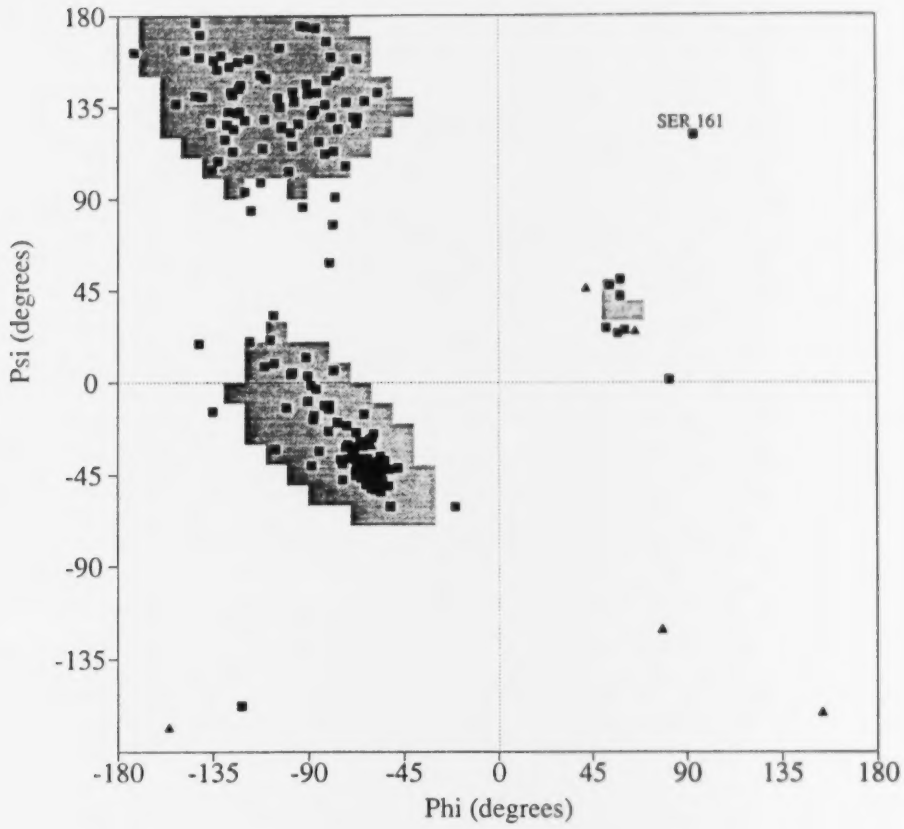


Figure 5.1. A Ramachandran plot of the refined NFP model after least squares cycle 93 drawn with PROCHECK (Laskowski *et al.*, 1993). The most favorable regions of the Ramachandran map are shaded. Glycine residues are marked with triangles and all other residues are marked with squares. Ser161 is at the end of a disordered loop.

Table 5.5

Regions of the NFP molecule with poorly associated electron density

Position in Sequence	Description
Asp43	break in density between C^{α} and C
Val44	break in density between C^{α} and C
Glu59	breaks in main chain density
Gln71	poor side chain density
Gln101	break in density between C^{α} and C
Ile121	break in density between C^{α} and C
Asn123	break in density between C^{α} and C
Asn137	break in density between C^{α} and C
Thr139	break in density between C^{α} and C
Gln157-Ser161	no side chain density and broken main chain density
Glu183	poor side chain density
Phe198	break in density between C^{α} and C
Asn228	weak density, disordered

There are two amino-acid residues that are not in agreement with the published amino acid sequence for *P. leiognathi* NFP (Illarionov *et al.*, 1990). Ile74 appears to be a valine, and Gly123 appears to be an asparagine or aspartic acid. We used traditional omit maps where we omitted the side chain atoms from the residue in question for several cycles of refinement and inspected the resultant difference maps when determining the possible identity of these amino acids. We also analyzed Bhat OMIT maps (Bhat, 1988). A Bhat OMIT map for Val74 is shown in Figure 5.2. Refined electron density for each of the flavin adducts is shown in Figures 5.3 and 5.4.

RESULTS

The β-Barrel

Two roughly orthogonal views of the refined NFP monomer are displayed in Figure 5.5. The views were chosen to accent the similarity of the central 7-stranded β-barrel to the canonical $(\beta\alpha)_8$ barrels of glycolate oxidase and other enzymes (Lesk *et al.*, 1989; Chapter 4). The NFP barrel has only seven strands and there is a large gap between strands β_2 and β_3, with enough room present for a missing eighth strand to be placed. This has prompted us to suggest that NFP evolved from a protein that contained an intact $(\beta\alpha)_8$ fold (Chapter 4). The packing of hydrophobic amino acids in the barrel interior of NFP, and its relation to the barrel cores of other $(\beta\alpha)_8$ proteins, notably glycolate oxidase (Lindqvist, 1989), has been analyzed in detail (Chapter 4). An interesting feature of the barrel is that strand β_6 extends from the barrel's N-terminal opening to form a platform with another short parallel β-strand (β_{6a}). This platform supports helix α_4 on the top of the barrel.

Hydrogen bonding interactions in the central seven-stranded β-barrel are given in Table 5.6 and depicted in Figure 5.6. The missing eighth strand of the barrel leaves a large cavity at the NFP dimer interface. This cavity is mostly filled with water molecules and amino acid side chains from the adjacent monomer. Several water molecules make hydrogen bonds with main chain atoms from strands β_2 and β_3 (Figure 5.6). The third strand (β_3) runs antiparallel to the others in the barrel, forming a β-hairpin with strand β_4. The presence of an antiparallel strand and the gap between strands β_2 and β_3 distorts the barrel geometry at the NFP dimer interface (Chapter 4) giving it an elliptical appearance (Figures 5.5a and 5.6).

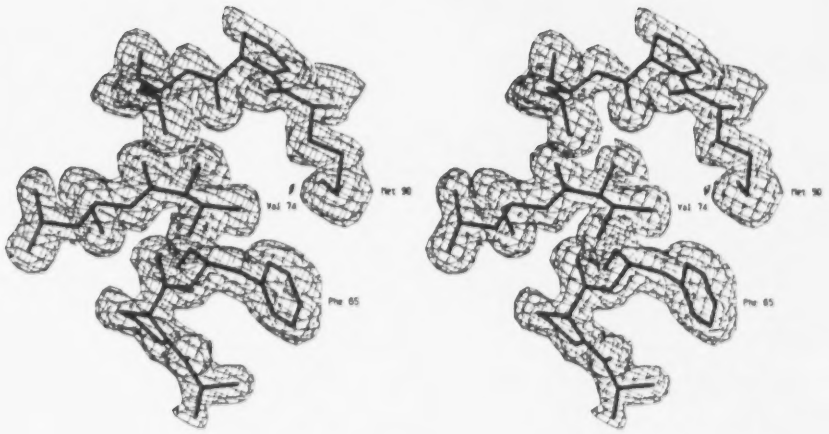
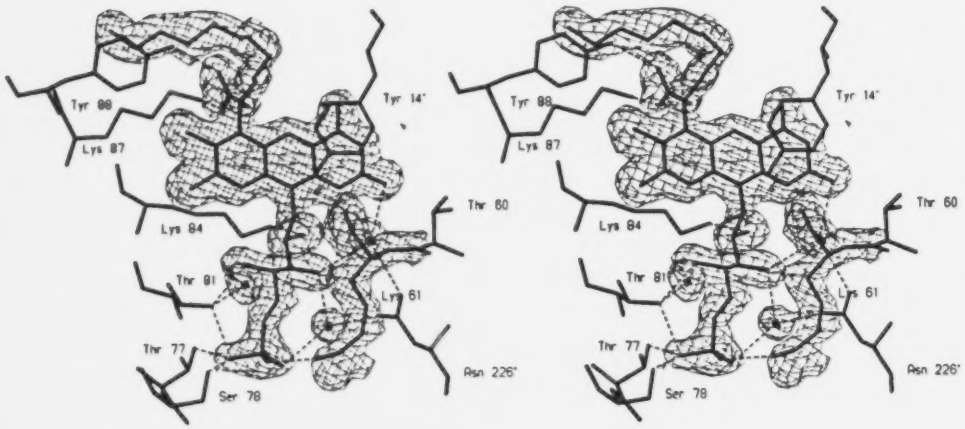
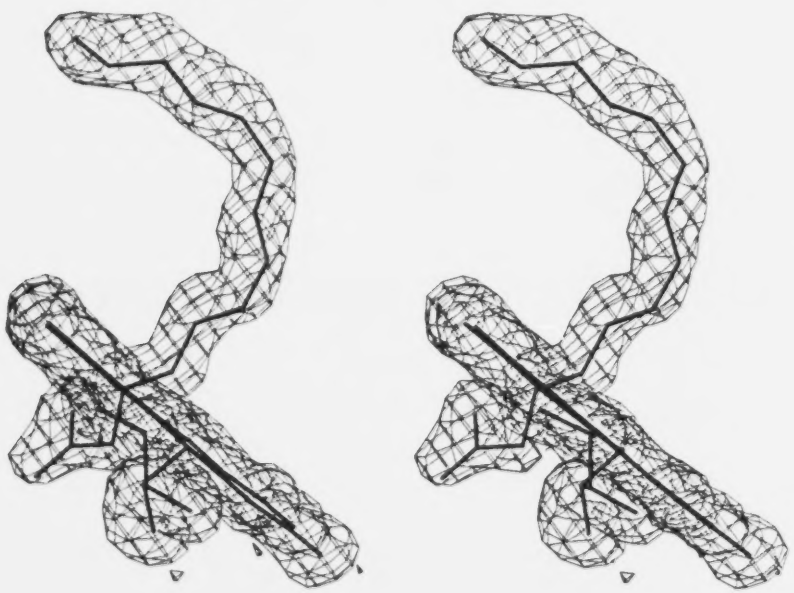


Figure 5.2. A stereo diagram of a portion of a Bhat OMIT map calculated in the vicinity of Val74 (Bhat, 1988). The refined model is drawn in dark lines and the map, contoured at 0.35 e/Å^3, is in thin chicken wire. Residues Phe65 Val74 and Met90 are labeled for clarity.

Figure 5.3. Stereo views of portions of the final $2|F_o|-|F_c|$ electron density map in the vicinity of the flavin cofactor at the dimer interface. The map is contoured at 0.35 e/Å^3. The cofactor and polypeptide models are drawn as thick lines. The map is in thin chicken wire. Maps are displayed with the BLOB option in Frodo which draws electron density within a 1.6Å radius of a defined coordinate set.

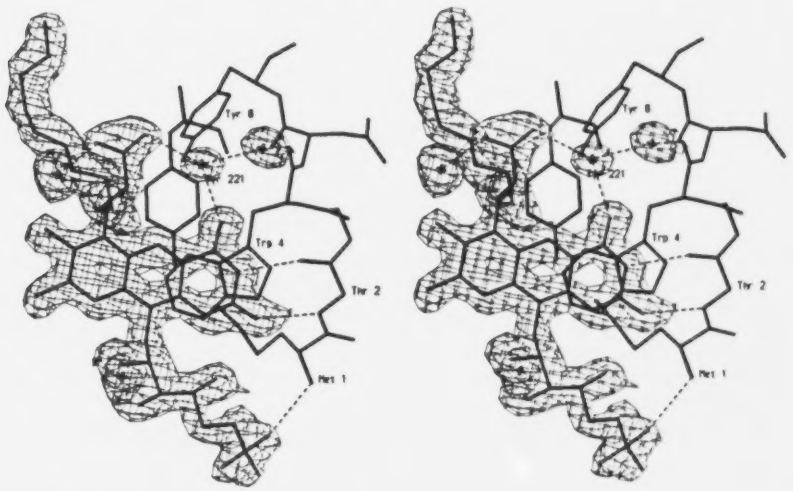


(a) A view of the isoalloxazine ring and ribityl chain of the FMN. Potential hydrogen bonds are drawn as dashed lines.

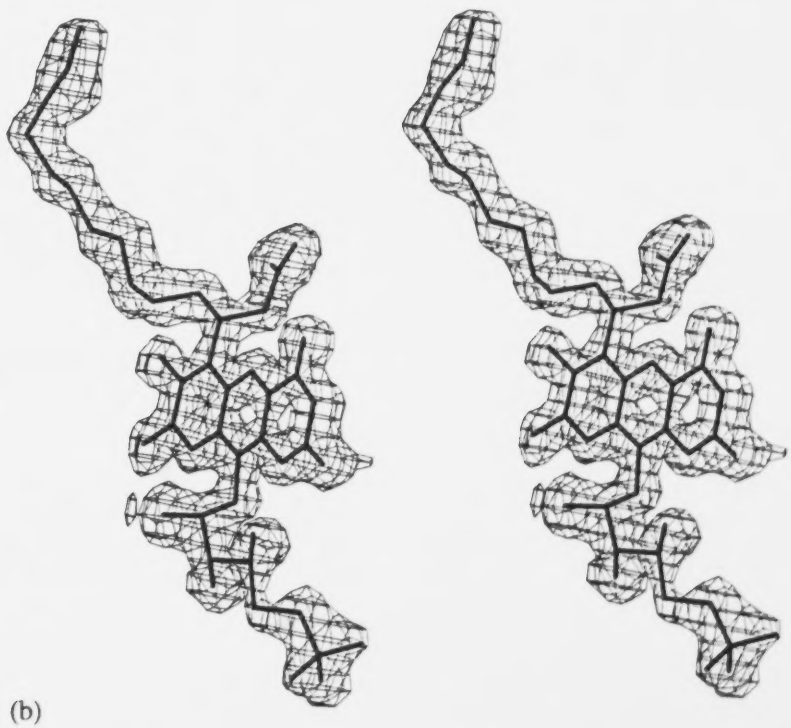


(b) A view of the myristic acid.

Figure 5.4. View of the myristlyated flavin at the C-terminus of the molecule. Drawn as in Figure 5.3.



(a)



(b)

Figure 5.5. Two Molscript (Kraulis, 1991) renditions of the NFP monomer, shown to highlight the similarity of NFP with $(\beta\alpha)_8$ barrel proteins.

(a) A view looking approximately down the barrel axis from the barrel's N-terminus.

(b) A view looking at the side of the NFP barrel. The dimer interface is in the foreground.

Table 5.6

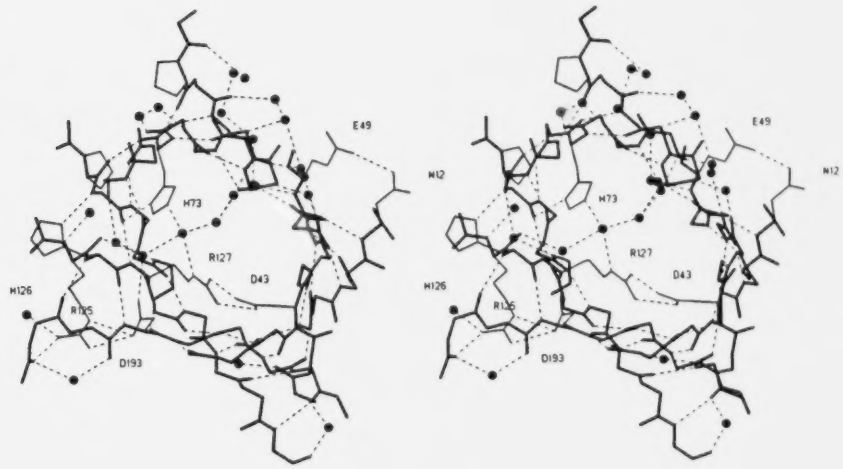
Hydrogen bonds associated with the β-sheet in NFP

Donor	Acceptor	Distance (Å)
Phe194, N	Asn5, O	2.7
Gly7, N	Phe194, O	2.9
Tyr6, O	Asp43, N	3.0
Phe196, N	Gly7, O	3.0
Val8, N	Val44, O	3.0
Ala46, N	Val8, O	3.1
Phe9, N	Phe196, O	2.9
Phe10, N	Ala46, O	3.1
Ser48, N	Phe10, O	2.8
Asn12, N	Ser48, O	3.0
Lys54, O	Leu62, N	2.7
Tyr56, N	Thr60, O	2.8
Glu59, N	Tyr56, O	3.2
Leu75, N	Pro64, O	2.9
Val66, N	His73, O	2.8
His73, N	Val66, O	2.7
Leu68, N	Gln71, O	2.7
Gln71, N	Leu68, O	3.2
Val74, N	Pro91, O	2.9
Leu93, N	Val74, O	2.9
Ala76, N	Leu93, O	2.8
Arg127, N	Leu92, O	2.9
Phe94, N	Arg127, O	2.9
Met129, N	Phe94, O	3.1

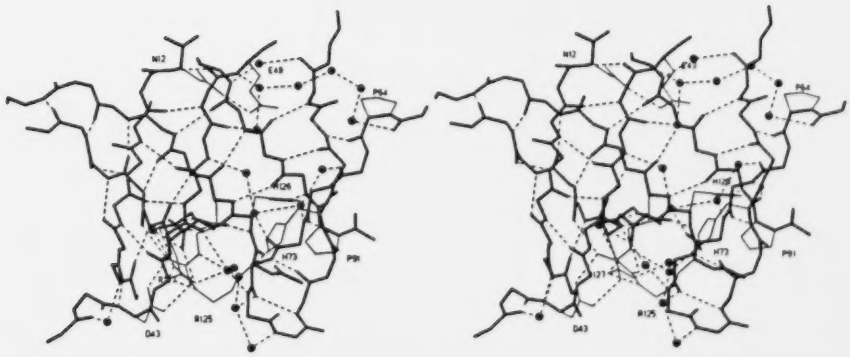
Table 5.6/contd.

His126, N	Lys191, O	2.8
Asp193, N	His126, O	2.6
Leu128, N	Asp193, O	2.8
Leu195, N	Leu128, O	2.9
Leu130, N	Leu195, O	2.9
Cys197, N	Leu130, O	2.9
Ala170, N	Phe131, O	3.0
Val132, N	Cys197, O	2.9
Asn133, N	Ala170, O	2.8
Gly172, N	Asn133, O	3.0
Asn135, N	Gly172, O	2.8

Figure 5.6. Two stereo views of the hydrogen bonding in the NFP β-barrel. Main chain atoms are drawn in thick and side chain atoms are thin. Hydrogen bonds are shown as dashed lines. Solvents are included as solid discs.



(a) A view looking from the C-terminus of the barrel.



(b) A side view of the barrel looking from the dimer interface. The C-terminus of the barrel is at the top of the figure.

Several charged amino acids are at the N-terminal opening of the β-barrel. Asp 43 and Arg127 make a bidentate salt bridge. The side chains of His73 and Arg127 donate hydrogen bonds to a water molecule. This water is part of a string of waters that spans the bottom or N-terminal opening of the barrel. At the edge of the top of the barrel, Arg125 makes a salt bridge with Asp193. At the top or C-terminal opening of the barrel, the side chains of Asn12 and Glu49 make a strong salt bridge. The start and end points of the β-strands are unchanged from our initial 2.3 Å determination of NFP (Moore *et al.*, 1993) (Chapter 3). Of note, however, is that water molecules often make bridging hydrogen bonds between main chain atoms on adjacent β-strands after the last β-sheet hydrogen bond.

α-Helices

There are seven α-helices in the NFP structure and five of these contain classical C-cap boxes (Schellman, 1980; Baker & Hubbard, 1984; Richardson & Richardson, 1988). The C-cap for helix α_7 at the C-terminus of NFP is shown in Figure 5.7. These capping interactions are characterized by a 3_{10} turn between residues i-1 and i-4 in the six residue box. The residue at i-3 is generally in the last amino acid in a truly α-helical conformation. Residue (i-2) is typically in a 3_{10} conformation ($\phi = -95$, $\psi = 7$). The residue at i-1 of the box is generally in a left-handed conformation ($\phi = 60$, $\psi = 45$) and is often asparagine or glycine. The amide nitrogen of residue i generally makes a hydrogen bond with the carbonyl oxygen of residue i-5. Residue i generally has a phi angle somewhere between -80 and -100. The i, i-5 hydrogen bonds of the capping boxes are summarized in Table 5.8 and the 3_{10} hydrogen bonds are listed with other reverse turns in the NFP structure (Table 5.9). Other helix capping interactions in NFP have been treated in detail in Chapter 4.

There are a number of other intricate hydrogen-bonding schemes in NFP that are associated with α-helices. Figure 5.8 shows one such set of interactions associated with adjacent helices α_1 and α_7. The hydrogen bonding potential of Arg33 on helix α_1 is fully satisfied by the carboxylate groups of Glu30, Glu37, and Asp38 similar to electrostatic interactions stabilizing the linker helix of muscle troponin C (Sundaralingam *et al.*, 1985) (Figure 5.8). All four residues are well ordered and exhibit exceptional electron density for polar residues on the surface of a helix. The N^{ζ} of Lys207 on nearby helix α_7 makes hydrogen bonds with the carbonyl oxygen of Asn27 and the $O^{\gamma 2}$ of Thr31. The side chain amide group of Asn27 makes a beautiful stacking interaction

Table 5.7

α-Helices in NFP

Helix	Residue range	α-Helical H-bonds
α1	Pro21-Glu37	14
α2	Pro79-Lys87	6
α3	Gln100-Lys116	15
α4	Pro138-Tyr155	12
α5	Ile162-Asn167	4
α6	Phe174-Gly187	12
α7	Gln204-Glu224	19

Table 5.8

i, i-5 Hydrogen bonds at the C-termini of α-helices

Donor (i)	Acceptor (i-5)	Distance (Å)	Conformation at i-2, (Type I turn)	Conformation at i-1, (3_{10} donor, α_L)	i
Met90, N	Ala85, O	3.0	ϕ = -106, ψ = 8	ϕ = 65, ψ = 28,	ϕ = -84
Val119, N	Ala114, O	2.8	ϕ = -86, ψ = -6	ϕ = 62, ψ = 49,	ϕ = -103
Ala158, N	Leu153, O	3.0	ϕ = -81, ψ = -6	ϕ = 50, ψ = 49,	ϕ = -61
Asn190, N	Ala185, O	2.7	ϕ = -106, ψ = 33	ϕ = 57, ψ = 24,	ϕ = 59
Leu227, N	Arg222, O	2.7	ϕ = -97, ψ = 5	ϕ = 59, ψ = 44,	ϕ = -93

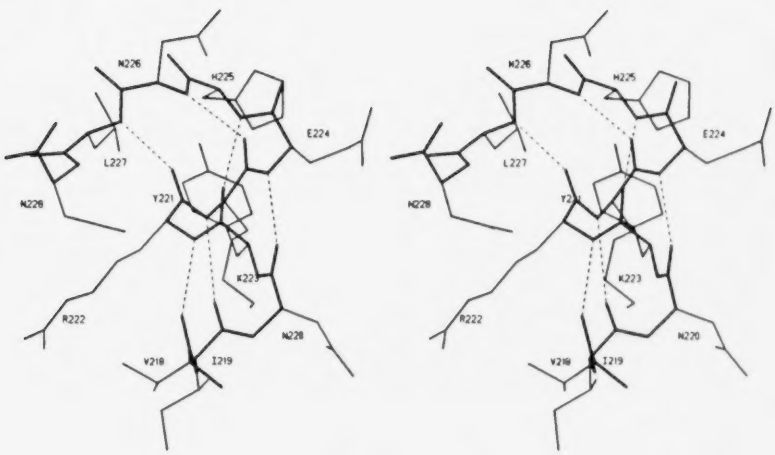


Figure 5.7. A stereo view of the C-capping box of helix α_7. Main chain atoms are drawn in thick and side chain atoms in thin. Hydrogen bonds are indicated with dashed lines.

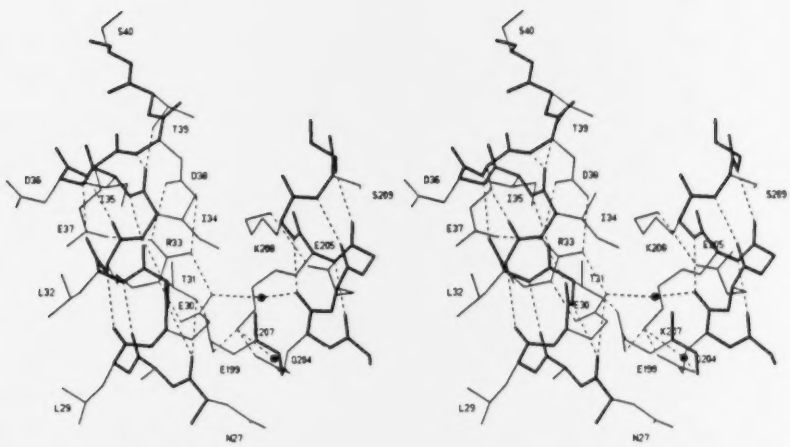


Figure 5.8. A stereo diagram showing the four residue salt bridge on helix α_1 and other interactions with helix α_7. Drawn as in Figure 5.7.

Table 5.9

Reverse (3^{10}) turns in the NFP structure

Donor	Acceptor	Distance (Å)	Comment	Turn Class	ϕ_2, ψ_2	ϕ_3, ψ_3
Gln19, N	Val16, O	2.9	Turn before helix	III'	43, 46	52, 28
Ile52, N	Glu49, O	3.3	Turn at end of β sheet	I	-79, 6	-81, -12
Glu59, N	Tyr56, O	3.2	Hairpin	I	-65, -30	-93, 12
Gln71, N	Leu68, O	3.1	Hairpin	II'	76, -118	-110, 5
Gly89, N*	Ala86, O	3.0	Helix C-cap	I	-54, -37	-106, 8
Asp98, N	Lys95, O	3.4	Turn before helix	I	-71, -24	-90, 4
Gln118, N*	Ala115, O	3.3	Helix C-cap	I	-62, -42	-86, -6
Ala150, N	Ile147, O	3.5	Helix kink	I	-56, -46	-84, -11
Gln157, N*	Ser154, O	3.4	Helix C-cap	I	-65, -45	-81, -6
Asn169, N	Ile166, O	3.0	Helix C-cap	I	-61, -30	-88, 0
Asn189, N*	Gln186, O	3.0	Helix C-cap	I	-59, -30	-106, 33
Lys191, N	Leu188, O	3.1	Turn at end of helix	III'	57, 23	59, 28
Met201, N	Phe198, O	3.1	Turn at end of β sheet	I	-63, -16	-79, -15
Asn226, N*	Lys223, O	3.2	Helix C-cap	I	-61, -39	-97, 5

* Reverse turns that are also part of helix C-capping boxes.

(d = 3.3 Å) with the side chain amide group of Gln 204. At the other end of helix α_1, Thr39,$O^{\gamma 2}$ and Asp38,N make hydrogen bonds to the carbonyl oxygen of Ile 34. On helix α_7, the $O^{\varepsilon 1}$ atom of Gln204 makes a hydrogen bond to the N^{ζ} of Lys207. There is an additional salt bridge between Lys208 and Glu205. The O^{γ} of Ser209 is hydrogen bonded to the carbonyl oxygen of Glu205.

Helix α_2 makes a number of polar interactions with the ribityl group of the FMN adduct at the NFP dimer interface but these have been described elsewhere (Moore *et al*, 1993; Chapter 3) and remain unchanged in the high resolution refinement. The N-terminus of helix α_3 interacts with the preceding reverse turn involving residues Lys95 and Asp98 and also with nearby helix α_5 (Figure 5.9). The guanidinium group of Arg103 makes hydrogen bonds with the carbonyl oxygens of Lys95 and Trp96 on the reverse turn and donates two hydrogen bonds to the carbonyl oxygen of Ser168 at the C-terminus of helix α_5. Lys102, also at the start of helix α_3, makes a good salt bridge with the carboxylate of Asp98. Having two positively charged residues in the first turn of an α-helix is statistically unfavoured (Richardson & Richardson, 1988) and would be thought to destabilize the helix dipole moment. However in the case of NFP nearby negatively charged groups and a number of good hydrogen bond acceptors seem to negate any unfavorable interactions with the helix dipole moment. This highlights the importance of taking into account specific local interactions when looking at amino acid preferences at helix ends and their relevance to helix stability. Looking at other interactions involving helices α_3 and α_5, the O^{γ} of Ser168 makes an interhelix capping hydrogen bond to the amide nitrogen of Gln100 on helix α_3. Asn169,$N^{\delta 2}$ donates hydrogen bonds to the carbonyl oxygens of Trp96 and Ile165. On helix α_3, the O^{γ} of Ser99 makes an N-cap hydrogen bond to the amide nitrogen of Lys102.

Helix α_5 also interacts with helix α_4 but in quite a different fashion. Helix α_4 follows a curved path and wraps around the hydrophobic surface of helix α_5 (Figure 5.10). The two helices are roughly perpendicular. The curvature of helix α_4 is facilitated by water molecules 261 and 317 which interrupt the normal α-helical hydrogen-bonding pattern between residues Asp151 and Ile147 and residues Tyr152 and Tyr148. This same effect is seen in helix B of chicken troponin C (Satyshur *et al.*, 1988). In NFP helix α_4 changes direction at Leu149 (which is in a 3_{10} conformation) and the amide nitrogen of Ala150 makes a 3_{10} hydrogen bond to the carbonyl oxygen of Ile147. A reverse turn that precedes helix α_1 (Val16 to Gln19) is in close proximity to the bend point of helix α_4 and the amide nitrogen of Gly17 donates a hydrogen bond to

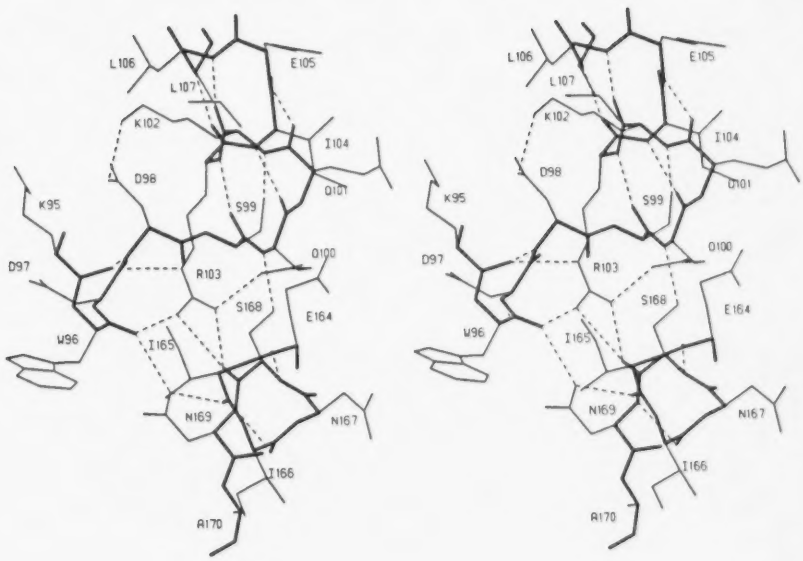


Figure 5.9. A stereo diagram illustrating interactions between the beginning of helix α_3, the end of helix α_5 and the 3_{10} turn prior to helix α_3. Drawn as in Figure 5.7.

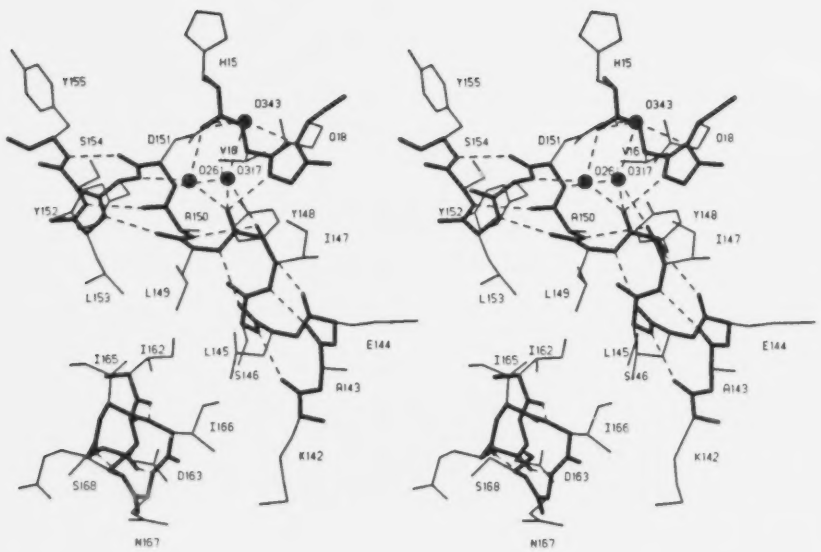


Figure 5.10. A stereo view of the kink in helix α_4 and its interactions with helix α_5 and the 3_{10} turn between His15 and Gln18. The helix kink is centered at residue Leu149 which is in a 3_{10} helical conformation. Drawn as Figure 5.7.

the carbonyl oxygen of Tyr149. The carbonyl oxygen of His15 makes a hydrogen bond with Wat261. A glycine at position 17 seems to be required in NFP to allow the close approach and hydrogen bonding of this reverse turn to the kink in helix α_4.

Reverse turns in the NFP structure

There are fourteen reverse turns in NFP (Table 5.9); eleven are of the type I variety (Crawford *et al.*, 1973), the other three are type III' and II'. Several of the turns have already been mentioned in reference to interactions with α-helices. One of the type I turns is the first β-hairpin at the dimer interface and involves strands β_{2a} and β_{2b} (Moore *et al.*, 1993; Chapter 3). This turn also contains an Asx turn (Figure 5.11). Apart from the reverse turn, this hairpin contains only two other main chain hydrogen bonds. The second hairpin is a type II' turn and connects strands β_3 and β_4 (Figure 5.12). There are four hydrogen bonds on this hairpin in addition to the reverse turn. At the beginning of strand β_3, the carbonyl oxygen of Ala63 makes an i, i+2 hydrogen bond with the amide nitrogen of Phe65. A string of four water molecules sits above and between the strands of the hairpin making hydrogen bonds with main chain atoms (Figure 5.12). The side chain of Lys70, at position i+2 of the type II' reverse turn makes two hydrogen bonds across the dimer interface to carbonyl groups at the C-terminus of helix α_1. The reverse turns involving Glu49-Ile52 and Lys95-Asp98 are both at the C-termini of β-strands and interact via their side chain atoms to make a beautiful network of hydrogen bonds (Figure 5.13). The phenolic OH of Tyr152 is also involved, making hydrogen bonds to Asp97,$O^{\delta 2}$ and Wat242. Wat242 makes hydrogen bonds to His51,$N^{\delta 1}$ and Ser48,O^{γ}. The carboxyl group of Asp97 makes a bifurcated salt bridge hydrogen bond to His50,$N^{\varepsilon 2}$ and makes a planar stacking interaction with the indole ring of Trp96 (d=3.3 Å). Trp96,$N^{\varepsilon 1}$ completes the ring of interactions with a hydrogen bond to Ser48,O^{γ}. Remember that the other end of the Lys95-Asp98 reverse turn makes several hydrogen bonds to the beginning of helix α_3. As mentioned before, several of the reverse turns are actually part of helix C-cap boxes (Tables 5.8 and 5.9).

Asx turns in NFP

There are five Asx turns in NFP and four of the five involve Asx residues (Table 5.10). However one of the turns uses the S^{γ} atom of Cys197 as the hydrogen bond acceptor (Figure 5.14). This cysteine also makes a good hydrogen bond with a water molecule that is part of a buried cluster of four waters near the fatty acid binding

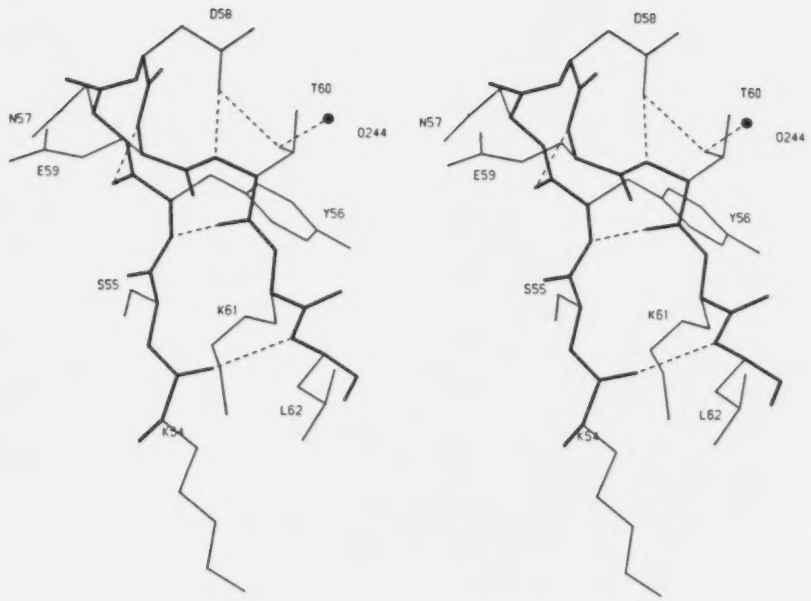


Figure 5.11. A stereo view of the combined β-hairpin and Asx turn centered at residue Asp58. Drawn as Figure 5.7.

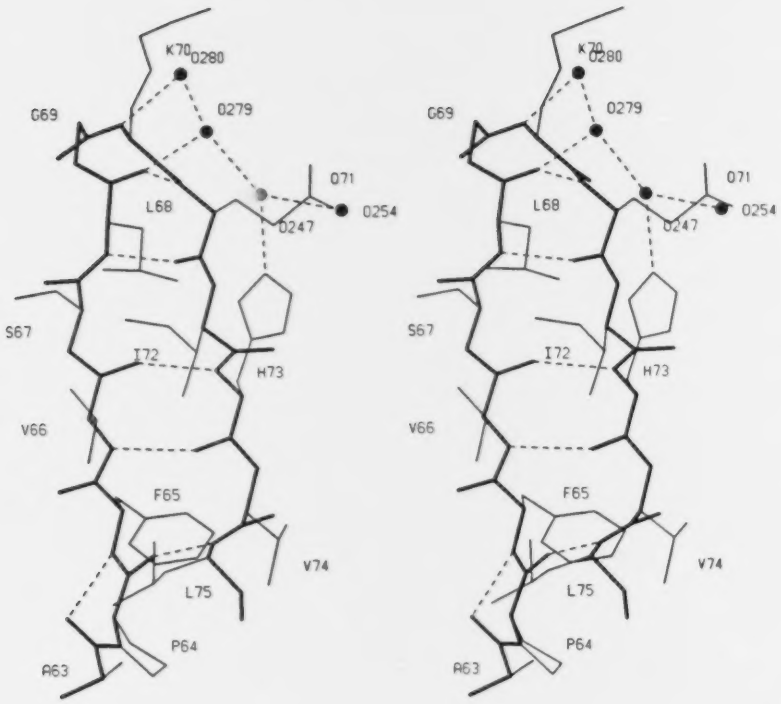


Figure 5.12. A stereo view of the β-hairpin connecting strands β_3 and β_4. Note the string of waters lying on "top" of the hairpin. Drawn as Figure 5.7.

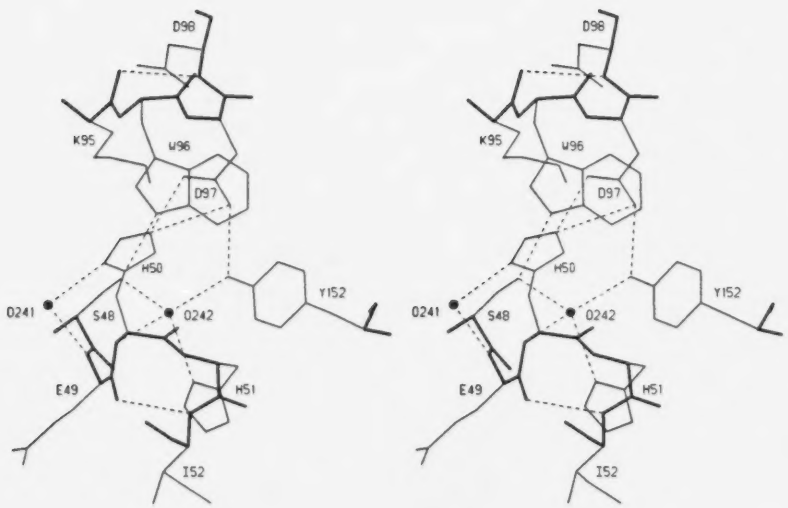


Figure 5.13. A stereo view of two 3_{10} turns that interact through a number of side chain hydrogen bonds. The turns are located at the C-termini of strands β_2 and β_3. Drawn as Figure 5.7.

Table 5.10

Asx turns in the NFP structure

Donor	Acceptor	Distance (Å)	Description	Residue	ϕ	ψ
Ala122, N	Asp120, $O^{\delta 1}$	3.1	Type I	Ile121	-116	20
Glu199, N	Cys197, S^{γ}	3.7	Type I	Phe198	-96	3
Glu60, N	Asp58, $O^{\delta 1}$	3.0	Type II	Glu59	81	4
Asn137, N	Asn135, $O^{\delta 1}$	2.9	Type III	Asp136	-57	-30
Thr139, N	Asn137, $O^{\delta 1}$	3.4	Type III	Pro138	-52	-42

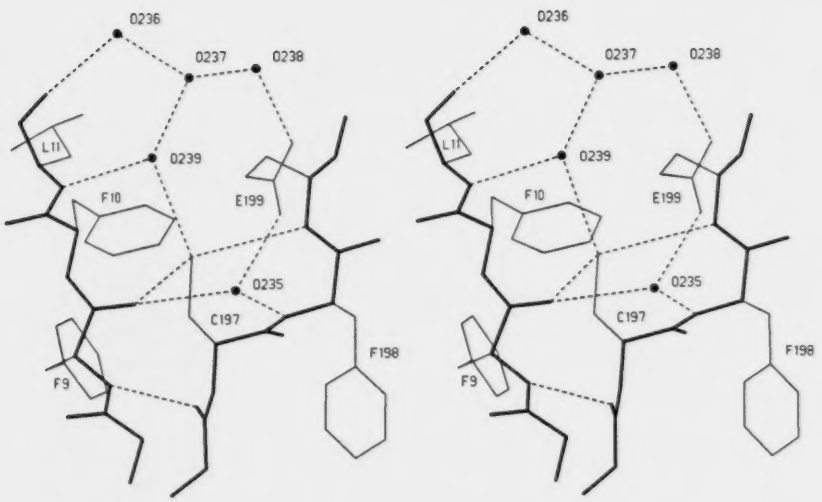


Figure 5.14. A stereo view of a rare Cys-containing Type I Asx turn at Cys197. Drawn as in Figure 5.7.

pocket at the dimer interface (Figure 5.15).

The flavin-fatty acid adducts

Electron density for each flavin adduct is shown in Figures 5.3 and 5.4. The absolute configuration at the fatty acid C3' methine atom is unambiguously the S enantiomer for both compounds. Torsion angles for both ribityl groups are given in Table 5.11 and the myristic acid torsion angles are listed in Table 5.12. Both fatty acid chains are bound in a curved conformation and this has been observed for fatty acids bound to other proteins (Weiss & Schulz, 1992; Benning *et al.*, 1992; Sacchettini *et al.*, 1989). The reason for this is that it is probably energetically less expensive to make a curved binding pocket versus an extended one in the interior of a protein. Most protein structural elements have some degree of curvature to them and the packing of amino acid side chains also is quite irregular.

The ribityl chains exhibit torsion angles that are very close to the conformational energetic minima of +/-60 or 180°. The majority of the myristic acid torsion angles are also close to energy minima although there are two cases of torsion angles near eclipsed conformations in each fatty acid (Table 5.12). The environment for the myristic acid at the dimer interface is shown in Figure 5.15. A cluster of five buried water molecules makes up part of the fatty acid binding cavity. The carbonyl oxygens of Leu11 and Phe13, the $O^{\delta 1}$ of Asn27, the hydroxyl of Tyr148, the S^{γ} of Cys197 and the $O^{\varepsilon 1}$ of Glu199 make good hydrogen bonds to these buried waters. The carboxylate of the myristic acid at the dimer interface binding site is coordinated by Lys87 and Tyr88 at the end of helix α_2. In fact these two residues are in a 3_{10} conformation, which allows the adjacent side chains to point nearly parallel to the plane of the carboxylate oxygens and make good hydrogen bonds. Human progastricsin and porcine pepsinogen utilize an identical LysTyr peptide on a 3_{10} helix to coordinate an aspartic acid residue at the active site, thus inhibiting these aspartic proteinase zymogens (James & Sielecki, 1986; James *et al*, 1994; Chapter 6). The isoalloxazine rings of both FMN moieties are relatively planar. The maximum deviation from the least squares plane is 0.028 Å for N10 at the C-terminal site. Both myristylated flavins are overlaid for comparison in Figure 5.16.

The dimer interface

One of the more conspicuous features of the NFP structure is the elaborate dimer interface. The molecular dyad is coincident with the crystallographic twofold symmetry

Table 5.11

Torsion angles for the FMN ribityl phosphate units

Torsion angle	FMN at dimer interface	FMN at C-terminus
$\chi 1$	101	-83
$\chi 2$	-177	152
$\chi 3$	-55	-63
$\chi 4$	-177	-177
$\chi 5$	62	-171
$\chi 6$	-179	161
$\chi 7$	-134	179

Torsion angles for the ribityl chains are defined as in Schulz *et al.* (1982) except that θ becomes $\chi 6$ and ϕ becomes $\chi 7$.

Table 5.12

Torsion angles for the myristic acid chains

Torsion angle	C14 fatty acid dimer interface	C14 fatty acid C-terminus
O1-C1-C2-C3	-45	67
C1-C2-C3-C4	108*	73
C2-C3-C4-C5	169	-180
C3-C4-C5-C6	179	-167
C4-C5-C6-C7	176	-82
C5-C6-C7-C8	95	-64
C6-C7-C8-C9	-172	-158
C7-C8-C9-C10	162	-161
C8-C9-C10-C11	-94	-176
C9-C10-C11-C12	177	-115*
C10-C11-C12-C13	-149	102*
C11-C12-C13-C14	156	165

*Nearly eclipsed conformations.

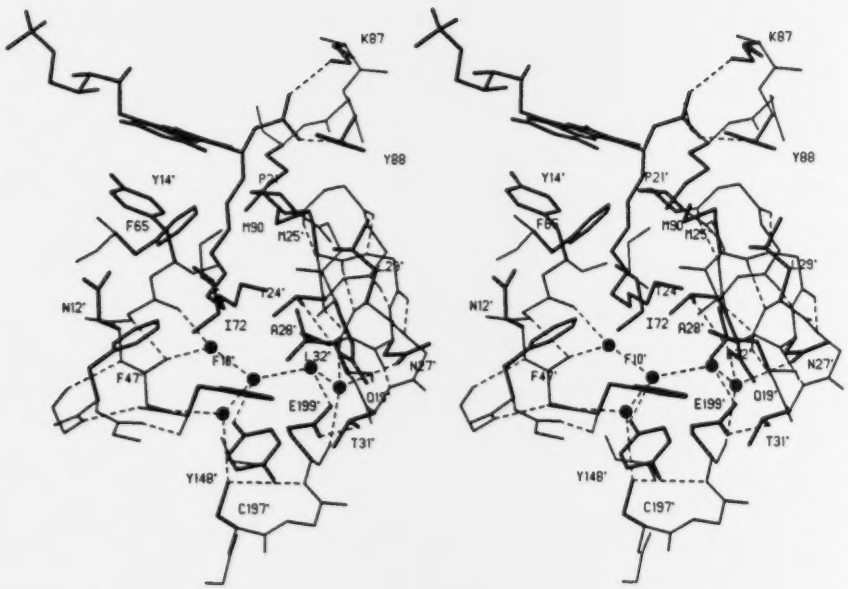


Figure 5.15. A stereo view of the myristic acid binding pocket for the myristylated FMN at the dimer interface. Amino acid residues from the other monomer are marked with a prime, e.g. F10'. Note that the majority of the residues in the fatty acid binding pocket come from one monomer. Almost all polar interactions with the flavin come from the other monomer. There is a cluster of five internal water molecules at the bottom of the pocket. Side chain bonds are thick, main chain bonds are thin. Waters are solid discs, and hydrogen bonds are dashed lines.

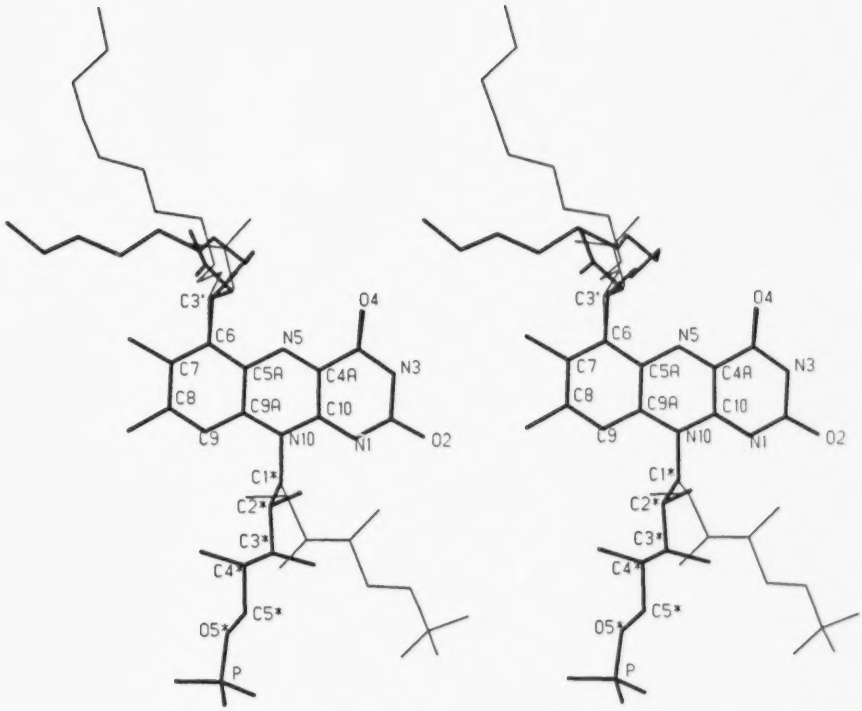


Figure 5.16. A stereo view highlighting the differences in conformation for the two myristylated flavins. The flavin from the dimer interface is thick, and the one from the C-terminus is thin.

axis at 0,y,1/4. A large solvent cavity is part of the dimer interface (Figure 5.16), it contains 34 water molecules (17 per monomer). The solvent cavity starts at the C-terminal opening of the β-barrel on one monomer, travels between strands β_2 and β_3 and exits out through the C-terminal opening of the β-barrel of the other monomer. This channel contains two charged residues from each monomer, Glu49 and Lys 54. They are depicted in Figure 5.17 and along with two water molecules make a six membered ring connected by six hydrogen bonds inside the cavity. Glu49, $O^{\varepsilon 1}$ also participates in a hydrogen bond with Wat321 which is hydrogen bonded to Phe65',O of the other monomer. (Symmetry related atoms are marked by a prime.)

A symmetrical lid-like structure sits over the top of the cavity and is made up of residues His51 to Asn57 from each monomer (Figure 5.17). These residues are hydrogen bonded to each other in an antiparallel fashion via a string of six water molecules that lie between the main chain atoms of each monomer segment (Figure 5.17). Tyr155,OH also participates in hydrogen bonds between these symmetry-related strands. Hydrogen bonds at the dimer interface are summarized in Table 5.13, the majority of these being mediated by water molecules that sit between the monomer polypeptide chains. At the bottom of the solvent cavity, near the N-terminal opening of the β-barrel, the hairpins connecting strands β_3 and β_4 from each monomer make a close approach and share hydrogen bonds with two (symmetry related) water molecules (Figures 5.18 and 5.19). Nearby, Val45,N makes a direct hydrogen bond across the interface to Gly69',O. Val45,O makes hydrogen bonds across the dimer interface to two alternate conformations of Ser67',O^{γ}. Val45,O is also hydrogen bonded to the aforementioned water that makes bridging hydrogen bonds between the symmetry related β_3-β_4 hairpins. Other residues involved in hydrogen bonds across the dimer interface are Asn12, Tyr14, Ala63, and Lys70.

There are a number of hydrophobic interactions at the dimer interface of NFP (Table 5.14 and Figure 5.20). Most of the hydrophobic residues come from helix α_1 and strands β_2 and β_3. The fatty acid chain is mostly buried in a pocket between helix α_1 and strand β_2 on one monomer (Figure 5.15) but does make van der Waals interactions with Phe65, Ile72 and Met90 from the other monomer. Direct van der Waals contacts between monomer side chains are made primarily by residues Asn12, Phe13, Tyr14, Met25, Leu32, Val45, Phe47, His51, Ile52, Lys54, Tyr56, Leu62, Ala 63, Phe65, Ile72, Tyr88, Met90, and Tyr155 (Table 5.14).

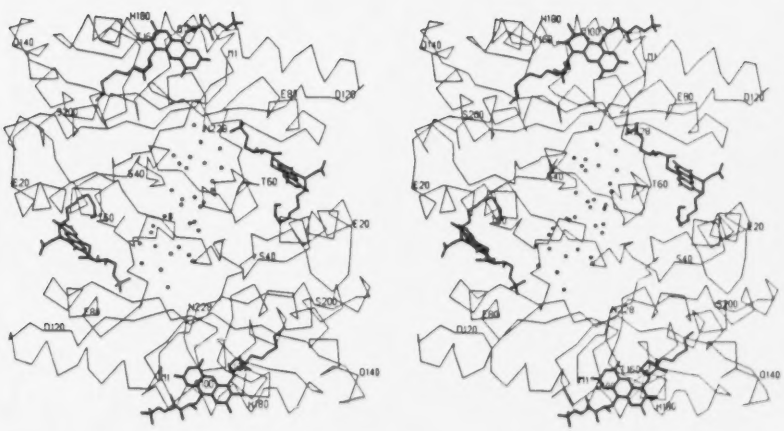


Figure 5.17. A stereo view of the solvent cavity at the dimer interface. The C^{α} tracing of the NFP polypeptide is drawn as thin lines, the myristylated flavins in thick, and water molecules as solid discs. The view looks approximately down the dimer interface at 0,y,1/4. The C-terminus of the barrel is in the foreground. Every twentieth residue is labeled.

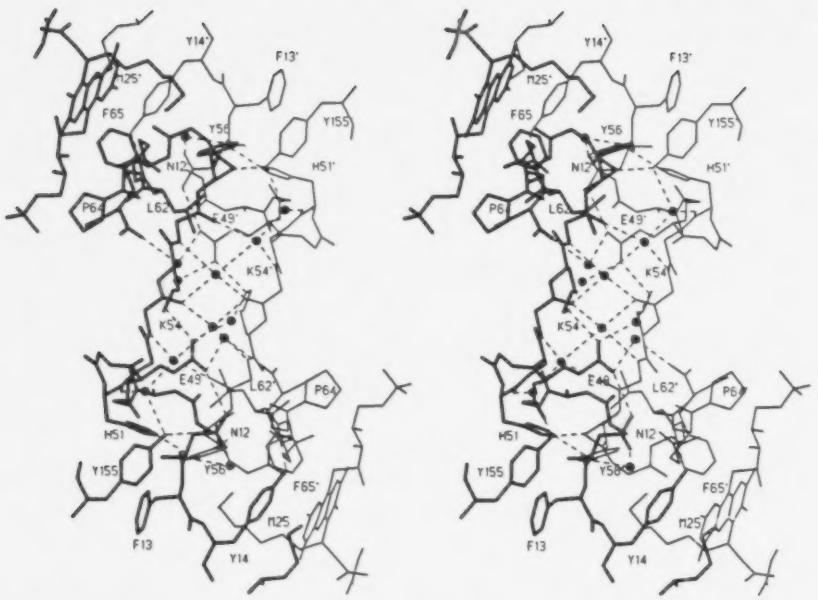


Figure 5.18. A stereo view of the solvent cavity's lid with antiparallel strands hydrogen bonded via a string of six bridging water molecules. One monomer is drawn in thick and the other monomer is drawn in thin and labeled with primes. Waters are drawn as solid discs and hydrogen bonds are shown as dashed lines. A six membered ring of hydrogen bonds that includes residues Glu49, Lys54, Glu49' and Lys54' lies just below the antiparallel lid.

Table 5.13

Hydrogen bonds at the dimer interface of NFP

Donor	Acceptor	Distance (Å)
Val45, N	Gly69', O	3.2
Gly69', N	Wat352, O	2.9
Wat352, O	Val45, O	3.5
Wat352, O	Gly69', O	2.9
Wat352, O	Wat352', O	2.8
Phe47, N	Wat277, O	2.8
Lys54, N^{ζ}	Glu49', $O^{\varepsilon 1}$	3.4
Lys54, N^{ζ}	Wat321, O	2.4
Wat321, O	Glu49', $O^{\varepsilon 1}$	2.8
Wat278, O	Glu49', $O^{\varepsilon 2}$	3.4
Ser67, $O^{\gamma a}$	Val45', O	3.0
Wat278, O	Phe65, O	2.9
Ser67, $O^{\gamma b}$	Val45', O	2.8
Asp53, N	Wat327, O	3.0
Wat276, O	Glu53', O	2.8
Wat276, O	Glu53, O	3.0
Ser55, N	Wat275, O	3.2
Wat327, O	Ser55',O	2.9
Wat327, O	Wat328', O	2.7
Wat328, O	His51, O	3.2
Tyr155', OH	Wat328', O	3.0
Asn57, N	Tyr155', OH	2.9

Table 5.13/contd.

Tyr56, OH	Asn12', O	2.9
His51, $N^{\varepsilon 2}$	Tyr56, OH	3.1
Wat245, O	Asn12', $O^{\delta 2}$	2.8
Wat245, O	Tyr56, OH	2.8
Tyr14', OH	Lys61, O	2.6
Asn12', $N^{\delta 2}$	Ala63, O	2.9
Tyr42, N	Wat240, O	3.0
Lys70', N^{ζ}	Wat240, O	3.5
Lys70', N^{ζ}	Ile35, O	3.2
Lys70', N^{ζ}	Wat375, O	2.6
Wat375, O	Thr39, O	3.3
Lys70', N^{ζ}	Wat374, O	2.4
Wat374, O	Ser40, O	2.6

Table 5.14

van der Waals contacts at the NFP dimer interface ≤ 4.0 Å

Monomer 1	Monomer 2	Distance (Å)
Met25, C^{ε}	Met90', S^{δ}	3.5
Met25, C^{ε}	Phe65' C^{ζ}	4.0
Met25, C^{γ}	Tyr88', $C^{\varepsilon 2}$	3.9
Met25, C^{γ}	Tyr88', C^{ζ}	3.7
Met25, C^{γ}	Tyr88', OH	3.8
Met25, C^{ε}	Tyr88', $C^{\delta 2}$	3.9
Leu32, $C^{\delta 1}$	Ile72', $C^{\gamma 1}$	3.8
Leu32, $C^{\delta 2}$	Ile72', $C^{\gamma 1}$	4.0
Val45, C^{β}	Ser67', $O^{\gamma a}$	3.6
Val45, $C^{\gamma 1}$	Ile72', $C^{\delta 1}$	3.4
Val45, O	Gly 69', O	3.9
Phe47, C^{ζ}	Phe65', C^{β}	3.4
Phe47 $C^{\varepsilon 1}$	Phe65', C^{β}	3.8
Phe47, $C^{\varepsilon 2}$	Phe65', C^{β}	3.8
Phe47, $C^{\varepsilon 2}$	Ser67', N	3.9
Phe47, $C^{\varepsilon 2}$	Ile72' $C^{\delta 1}$	3.7
Phe47, C^{ζ}	Ile72', $C^{\delta 1}$	3.7
Phe47, C^{ζ}	Ile72', $C^{\gamma 2}$	3.8
His51, $C^{\delta 2}$	Tyr56', $C^{\varepsilon 2}$	3.6
His51, $C^{\delta 2}$	Tyr56' C^{ζ}	4.0
His51, $C^{\delta 2}$	Tyr56' OH	3.8
His51, $N^{\varepsilon 2}$	Tyr56', $C^{\varepsilon 2}$	3.8
His51, $N^{\varepsilon 2}$	Tyr56', C^{ζ}	3.7
Ile52, $C^{\delta 1}$	Ser55', O	3.5

Table 5.14/contd.

Ile52, $C^{\delta 1}$	Tyr56', $C^{\delta 2}$	3.9
Ile52, $C^{\delta 1}$	Tyr56', $C^{\varepsilon 2}$	4.0
Ile52, $C^{\gamma 1}$	Leu62', $C^{\delta 1}$	3.8
Ile52, $C^{\gamma 2}$	Lys54', C^{α}	3.8
Ile52, $C^{\gamma 2}$	Lys54', C^{β}	3.9
Tyr56, $C^{\varepsilon 2}$	Phe13', $C^{\delta 1}$	3.7
Tyr56, $C^{\varepsilon 2}$	Phe13', $C^{\varepsilon 1}$	3.8
Tyr56, OH	Phe13' C^{α}	3.4
Tyr56, OH	Phe13', $C^{\delta 1}$	3.7
Tyr56, $C^{\varepsilon 1}$	Tyr14', $C^{\varepsilon 2}$	4.0
Tyr56, C^{α}	Tyr155', OH	3.4
Tyr56, C^{β}	Tyr155', OH	3.6
Tyr56, $C^{\delta 2}$	Tyr155', C^{ζ}	3.5
Tyr56, $C^{\delta 2}$	Tyr155', OH	3.5
Tyr56, C	Tyr155', OH	3.7
Lys61, O	Tyr14', $C^{\varepsilon 2}$	3.2
Lys61, O	Tyr14', C^{ζ}	3.3
Leu62, C^{γ}	Asn12', $N^{\delta 2}$	3.6
Leu62, $C^{\delta 1}$	Asn12', $N^{\delta 2}$	3.6
Leu62, $C^{\delta 1}$	Asn12', O	3.5
Leu62, C	Tyr14', OH	3.6
Leu62, $C^{\delta 2}$	Glu49', C^{δ}	3.7
Leu62, $C^{\delta 2}$	Glu49', $O^{\varepsilon 1}$	3.9
Leu62, $C^{\delta 2}$	Glu49', $O^{\varepsilon 2}$	3.9
Ala63, C^{β}	Asn12', $O^{\delta 1}$	3.8
Ala63, N	Tyr14', OH	3.1
Ala63, C^{β}	Tyr14', OH	3.2

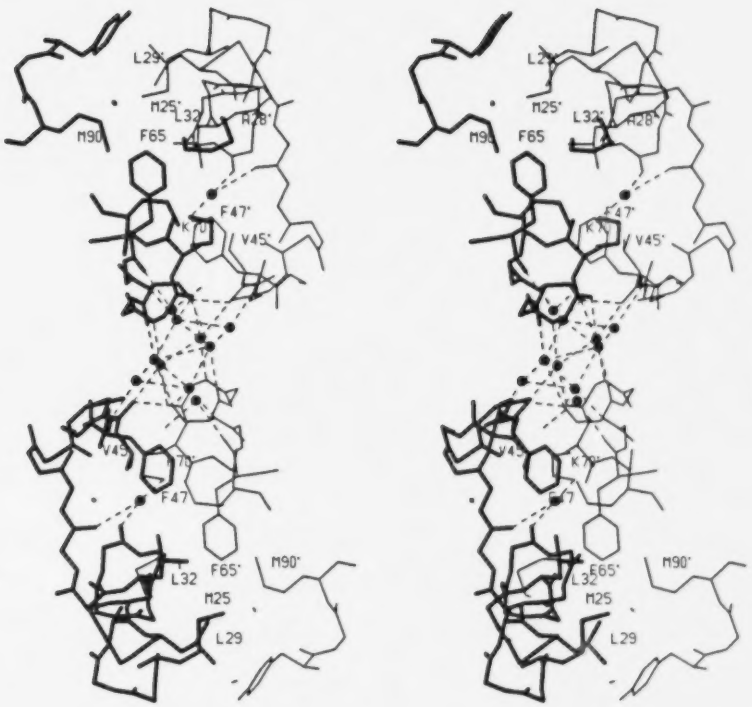


Figure 5.19. A stereo view of the bottom portion of the solvent cavity between the monomers highlighting hydrogen bonds between the β-hairpin between strands β_3 and β_4 and it's two-fold related mate. Drawn as in Figure 5.18.

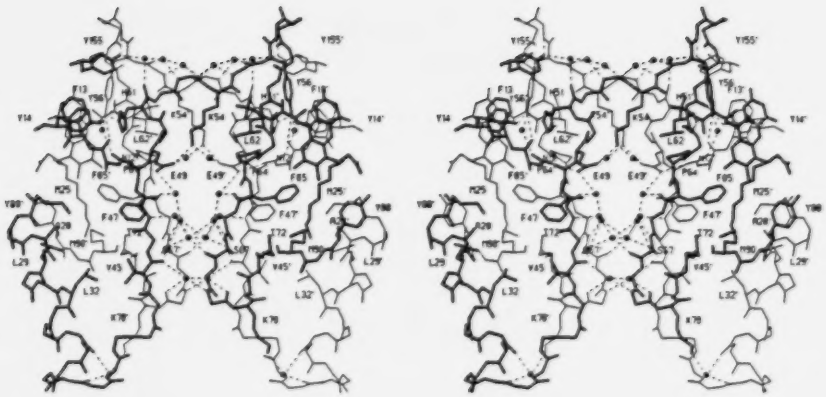


Figure 5.20. A side-on stereo view of the dimer interface of NFP, illustrating hydrogen bonds and van der Waals contacts between the monomers. The C-terminus of the β-barrel is at the top of the figure. Drawn as Figure 5.18.

A noteworthy van der Waals contact at the dimer interface is made from Tyr56,OH to Phe13',C^{α} (d = 3.4 Å) (Figure 5.21). If one assumes that the C^{α}-H is acidic enough to be a hydrogen bond donor, the geometry of this contact is ideal for the C^{α}-H to donate a hydrogen bond to the OH of Tyr56. The ^{1}H NMR spectra of amino acids shows the C^{α}-H proton to be highly deshielded, suggesting that the C^{α}-H bond is strongly polarized. Indeed C-H ··· O hydrogen bonds have been extensively documented in the literature (Taylor & Kennard, 1982). The hydrogen bond donating potential of the C^{α}-H is generally ignored by protein crystallographers, but it has been shown to be involved in energetically favourable interactions in glycine crystals (Berkovitch-Yellin & Leiserowitz, 1984).

Crystal packing

In the NFP unit cell, there are two layers of four NFP monomers packed in the ac plane at approximately 1/4 and 3/4 in b (Figure 5.22). Each ac layer contains an NFP dimer and two monomers. The monomers' mates are in adjacent unit cells. The major crystallographic contact is, of course, the dimer interface with nearly 3200 $Å^2$ of solvent accessible surface area buried (Moore *et al.*, 1993; Chapter 3). The primitive molecule at x,y,z has six nearest neighbors in the ac plane, four of them being associated with specific symmetry operations on the primitive molecule (Figure 5.23 and Table 5.15). Two of these contacts are just reverse screw motions of two of the others. There is only one molecular interaction parallel to the b axis and it involves six hydrogen bonds and several van der Waals contacts (Table 5.15 and Figure 5.24). These interactions are all that is holding adjacent ac layers together and yet is enough to give crystals that diffract to 1.6 Å resolution. One of the symmetry interactions in the ac plane involves a sulphate anion that sits on the 0,y,1/4 two-fold axis and sits between helix α_3 on adjacent NFP monomers. The $N^{\delta 1}$ atom of His109 on this helix makes a bifurcated hydrogen bond with two of the sulphate oxygens. This interaction is related across the twofold axis and bridges the two helices (Figure 5.25). Another salt bridge at this interface involves Glu105 and Lys116'.

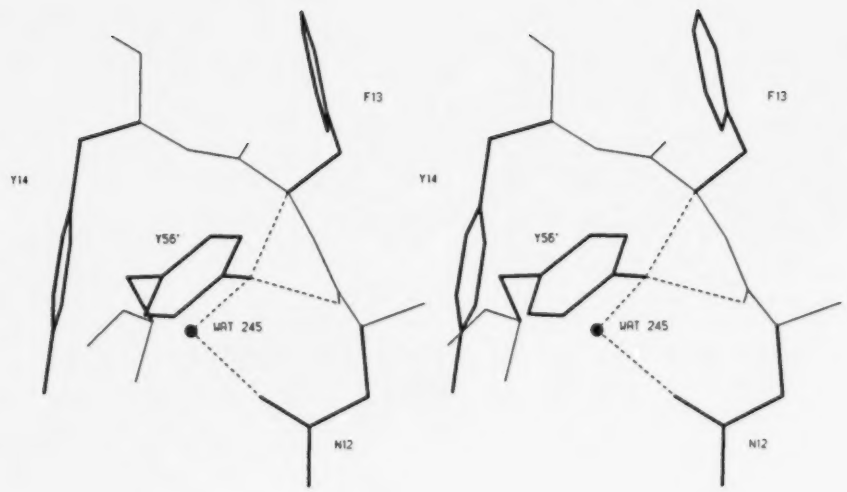


Figure 5.21. A potential hydrogen bond between Tyr56', OH and Phe13, C^{α}-H (d = 3.4 Å) at the dimer interface is show in stereo. Drawn as in Figure 5.15.

Table 5.15

Protein-protein contacts in NFP involving crystallographic symmetry

Molecule at x,y,z	Transformed molecule at x',y',z'	Transformation mapping x,y,z onto x',y',z'	Number of protein-protein contacts ≤ 4 Å
D136, N137, Q140	K87, Y88, N118, V119	1/2-x,1/2-y,1/2+z 2_1 axis at (1/4,1/4,z)	15
I104, E105, N108, H109 (Sul231)	A112, A115, K116, H109 (Sul231)	1-x,y,1/2-z 2-fold axis at (1/2,y,1/4)	13
		-x,y,1/2-z 2-fold axis at (0,y,1/4)	extensive
P138, N167, N173, D175, T176, H179	S26, E205, S22, S22, L23, S26, E20	1/2+x,1/2-y,1-z 2_1 axis at (x,1/4,1/2)	20
D58, E59, K61	M1, H225, N226	1/2-x,1/2+y,1/2-z 2_1 axis at (1/4,y,1/4)	23

There are three other nearest neighbors. A molecule at x-1/2,1/2-y,1-z makes the same contacts as the molecule at 1/2+x,1/2-y,1-z, but with the interacting atoms on the molecule at x,y,z now being on the transformed molecule and vice-versa. The symmetry transformation is also the same, only the transformation (here a screw motion parallel to x) changes direction. In the same way, the molecule at 1/2-x,y-1/2,1/2-z makes identical interactions as the transformation 1/2-x,1/2+y,1/2-z. So does the molecule at 1/2-x,1/2-y,z-1/2 to the one at 1/2-x,1/2-y,1/2+z. In total, there are six nearest neighbors in the ac plane and two neighbors along the b axis.

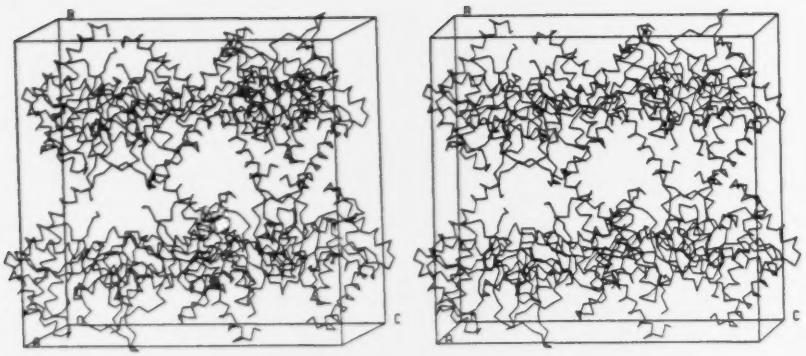


Figure 5.22. A stereo diagram showing the molecular packing in the NFP unit cell. NFP is represented as a C^{α} tracing and the unit cell is drawn in solid lines with the origin and axes labeled.

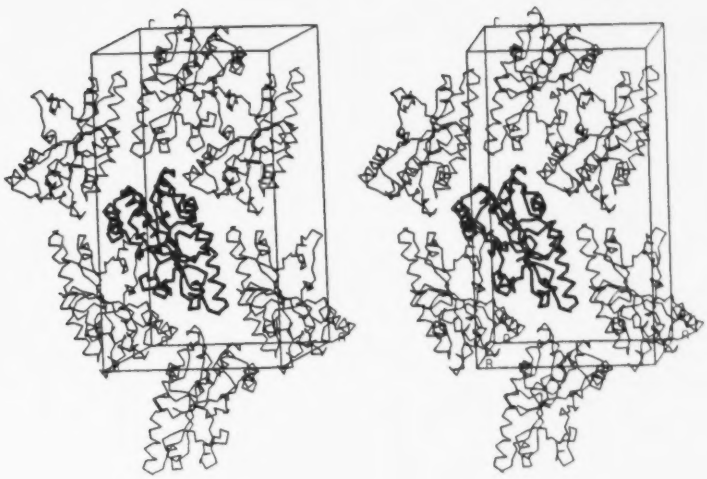


Figure 5.23. A stereo view of the molecular packing in the ac plane. The NFP monomer at x,y,z (in thick) has six nearest neighbors. Looking approximately down the b axis.

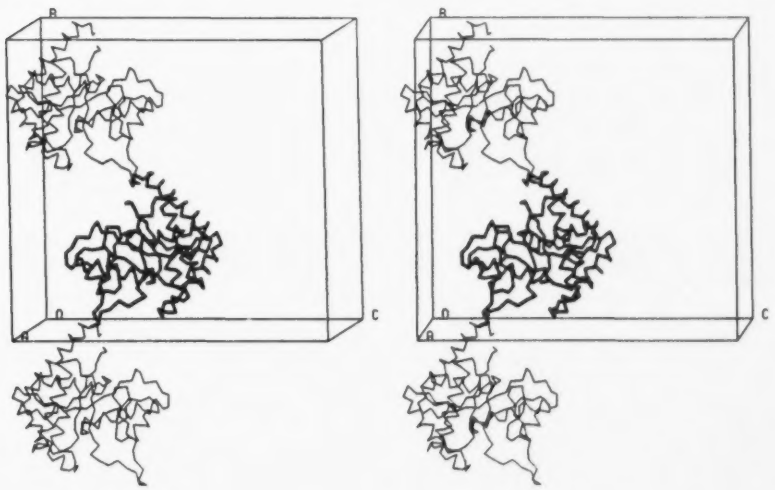


Figure 5.24. A stereo view of the molecular packing along the b axis. The molecule at x,y,z (in thick) has two nearest neighbors related by a 2_1 screw axis.

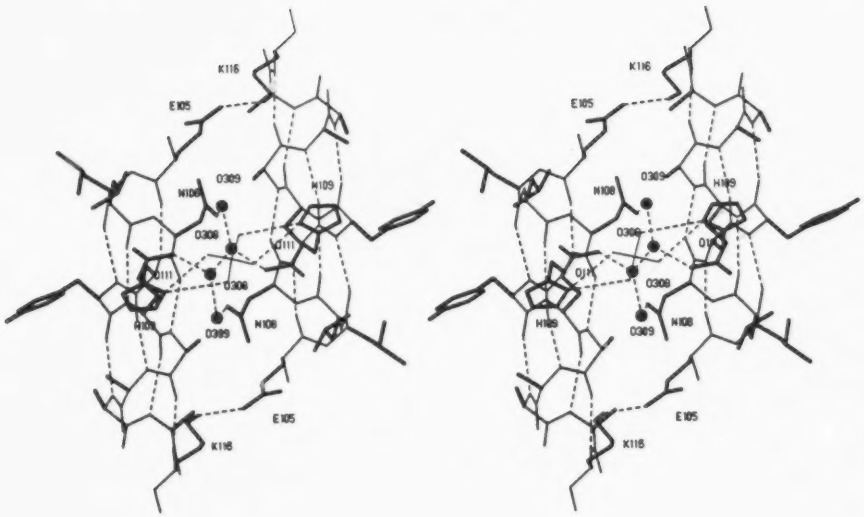


Figure 5.25. A stereo view of the sulphate that sits on the crystallographic two-fold axis at 1/2,y,1/4. Both H109, $N^{\delta 1}$ and H109', $N^{\delta 1}$ participate in bifurcated hydrogen bonds with two of the sulphate oxygens. Drawn as in Figure 5.15.

CONCLUSIONS

The structure of NFP has been refined to R=0.175 for data extending to 1.6 Å resolution. Hydrogen bonding is extensive and there are five C-cap boxes observed at the C-termini of α-helices. Both flavin cofactors are well ordered and have well-defined electron density for the FMN moieties and myristic acid chains, respectively. Many of the hydrogen bonds at the extensive dimer interface are mediated by well-ordered water molecules. The presence of only one enantiomer (S) of the flavin adduct, suggests that an enzyme participated in its formation.

REFERENCES

Baker, E. N. & Hubbard, R. E. (1984). Hydrogen bonding in globular proteins. *Prog. Biophys. Molec. Biol.* **44**, 97-179.

Benning, M. W., Smith, A. F., Wells, M. A. & Holden, H. M. (1992). Crystallization, structure determination and least squares refinement to 1.75 Å resolution of the fatty-acid-binding protein isolated from *Manduca sexta* L. *J. Mol. Biol.* **228**, 208-219.

Berkovitch-Yellin, Z. & Leiserowitz, L. (1984). The role played by C-H···O and C-H···N interactions in determining molecular packing and comformation. (1984) *Acta. Cryst.* **B40**, 159-165.

Bhat, T. N. (1988). Calculation of an OMIT map. *J. Appl. Cryst.* **21**, 279-281.

Crawford, J. L., Lipscomb, W. N. & Schellman, C. G. (1973). The reverse turn as a polypeptide conformation in globular proteins. *Proc. Natl. Acad. Sci. USA* **70**, 538-542.

Fields, B. A., Guss, J. M., Lawrence, M. C. & Nakagawa, A. (1992). The Weissenberg method for the collection of X-ray diffraction data from macromolecular crystals: modifications to the data-processing program WEIS. *J. Appl. Cryst.* **25**, 809-811.

Hamilton, W. C., Rollett, J. S. & Sparks, R. A. (1965). On the relative scaling of X-ray photographs. *Acta Cryst.* **18**, 129-130.

Hendrickson, W. A. & Konnert, J. H. (1980). Incorporation of stereochemical information into crystallographic refinement. In *Computing in Crystallography* (Diamond, R., Ramaseshan, S. & Venkatesan, K., eds.), pp. 13.01-13.23, Indian Academy of Sciences, International Union of Crystallography, Bangalore, India,

Higashi, T. (1989). The processing of diffraction data taken on a screenless Weissenberg camera for macromolecular crystallography. *J. Appl. Crystallogr.* **22**, 9-18.

Illarionov, B. A., Blinov, V. M., Donchenko, A. P., Protopopova, M. V., Karginov, V. A., Mertvetsov, N. P. & Gitelson, J. I. (1990). Isolation of bioluminescent functions from *Photobacterium leiognathi*: analysis of *lux*A, *lux*B, *lux*G and neighboring genes. *Gene* **86**, 89-94.

James, M. N. G. & Sielecki, A. R. (1986). Molecular structure of an aspartic proteinase zymogen, porcine pepsinogen, at 1.8 Å resolution. *Nature* **319**, 33-38.

James, M., Moore, S., Sielecki, A., Chernaia, M. & Tarasova, N. (1994). The molecular structure of human progastricsin and its comparison with that of porcine pepsinogen. In *Structure and Function of the Aspartic Proteinases* (Takahashi, K., ed.), Plenum Press, New York. In press.

Jones, T. A. (1978). A graphics model building and refinement system for macromolecules. *J. Appl. Cryst.* **11**, 268-272.

Kasai, S., Fujii, S., Miura, R., Odani, S., Nakaya, T. & Matsui, K. (1991). Structure of FP390 including its prosthetic group (Q-flavin):physiological significance of light emitting reaction in luminous bacteria. In *Flavins and Flavoproteins* (Curti, B., Ronchi, S. & Zanetti, G., eds.), pp. 285-288, Walter de Gruyter & Co., Berlin.

Kasai, S., Matsui, K. & Nakamura, T. (1982). Structure of P-Flavin from *P. phosphoreum*. In *Flavins and Flavoproteins* (Massey, M. & Williams, C. H., eds.), pp. 459-462, Elsevier, North Holland.

Kasai, S., Matsui, K. & Nakamura, T. (1987). Purification and some properties of FP390 from P. phosphoreum. In *Flavins and Flavoproteins* (Edmondson, D. E. & McCormick, D. B., eds.), pp. 647-650, Walter de Gruyter & Co., Berlin.

Kita, A., Kasai, N., Kasai, S., Nakaya, T. & Miki, K. (1991). Crystallization and preliminary X-ray diffraction studies of a flavoprotein, FP390, from a luminescent bacterium, *Photobacterium phosphoreum*. *J. Biochem. (Tokyo)* **110**, 748-750.

Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.* **24**, 946-950.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283-291.

Lesk, A. M., Brändén, C-I & Chothia, C. (1989). Structural principles of α/β barrel proteins: the packing of the interior of the sheet. *Proteins* **5**, 139-148.

Meighen, E. A. (1991). Molecular biology of bacterial bioluminescence. *Microbiol. Rev.* **55**, 123-142.

Moore, S. A., James, M. N. G., O'Kane, D. J. & Lee, J. (1992). Crystallization of *Photobacterium leiognathi* NFP, an unusual flavoprotein with limited sequence identity to bacterial luciferase. *J. Mol. Biol.* **224**, 523-526.

Moore, S. A., James, M. N. G., O'Kane, D. J. & Lee, J. (1993). Crystal structure of a flavoprotein related to the subunits of bacterial luciferase. *EMBO J.* **12**, 1767-1774.

O'Kane, D. J. & Prasher, D. C. (1992). Evolutionary origins of bacterial bioluminescence. *Mol. Microbiol.* **6**, 443-449.

O'Kane, D. J., Vervoort, J., Muller, F. & Lee, J. (1987). Purification and characterization of an unusual nonfluorescent flavoprotein from *Photobacterium leiognathi*. In *Flavins and Flavoproteins* (Edmondson, D. E. & McCormick, D. B., eds.), pp. 641-645, Walter de Gruyter & Co., Berlin.

Raibekas, A. A. (1991). Green flavoprotein from *P. leiognathi*: purification, characterization and identification as the product of the lux G(N) gene. *J.*

Read, R. J. (1986). Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta. Cryst.*. **A42**, 142-149.

Richardson, J. S. & Richardson, D. C. (1988). Amino acid preferences for specific locations at the ends of α helices. *Science* **240**, 1648-1652.

Sacchettini, J. C., Gordon, J. I. & Banaszak, L. J. (1989). Crystal structure of rat intestinal fatty-acid-binding protein: refinement and analysis of the *Escherichia coli*-derived protein with bound palmitate. *J. Mol. Biol.* **208**, 327-339.

Sakabe, N. (1983). A focusing Weissenberg camera with multi-layer-line screens for macromolecular crystallography. *J. Appl. Cryst.* **16**, 542-547.

Sakabe, N. (1991). X-ray diffraction data collection system for modern protein crystallography with a Weissenberg camera and an imaging plate using synchrotron radiation. *Nucl. Instrum. Methods Phys. Res.* **Sect. A303**, 448-463.

Satyshur, K. A., Rao, S. T., Pyzalska, D., Drendel, W., Greaser, M. & Sundaralingam, M. (1988). Refined structure of chicken skeletal muscle troponin C in the two-calcium state at 2 Å resolution. *J. Biol. Chem.* **263**, 1628-1647.

Schellman, C. (1980). The α_L conformation at the ends of helices. In *Protein Folding* (R. Jaenicke, ed.), pp. 53-61, Elsevier, Amsterdam.

Schulz, G. E., Schirmer, R. H. & Pai, E. F. (1982). FAD-binding site of glutathione reductase. *J. Mol. Biol.* **160**, 287-308.

Soly, R. R., Mancini, J. A., Ferri, S. R., Boylan, M. & Meighen, E. A. (1988). A new *lux* gene in bioluminescent bacteria codes for a protein homologous to the bacterial luciferase subunits. *Biochem. Biophys. Res. Commun.* **155**, 351-358.

Sundaralingam, M., Drendel, W. & Greaser, M. (1985). Stabilization of the long central helix of troponin C by intrahelical salt bridges between charged amino acid side chains. *Proc. Natl. Acad. Sci. USA* **82**, 7944-7947.

Taylor, R. & Kennard, O. (1980). Crystallographic evidence for the existence of C-H···O, C-H···N, and C-H···Cl hydrogen bonds. *J. Am. Chem. Soc.* **104**, 5063-5070.

Thiessen, W. E. & Levy, H. A. (1973). International Union of Crystallography World List of Crystallographic Computer Programs. *J. Appl. Crystallogr.* **6**, 309-309.

Weiss, M. S. & Schulz, G. E. (1992). Structure of porin refined at 1.8 Å resolution. *J. Mol. Biol.* **227**, 493-509.

CHAPTER 6

CRYSTAL AND MOLECULAR STRUCTURES OF HUMAN PROGASTRICSIN AT 1.62 Å RESOLUTION[1]

INTRODUCTION

The aspartic proteinases (EC 3.4.23) comprise a large family of enzymes produced by both uni- and multicellular eukaryotic organisms. These enzymes are characterized by two catalytic aspartate residues at the active site, and generally display optimal activity at acid pH (Foltmann, 1981). Mature aspartic proteinases are polypeptides of about 320 amino acid residues having a molecular pseudo-two-fold axis of symmetry that passes through the active site between the catalytic aspartates. The similarity of the folds of the N- and C-terminal domains has suggested the hypothesis that modern cellularly derived aspartic proteinases evolved from a symmetrical dimer that underwent gene duplication and fusion events (Tang *et al.*, 1978; Hobart *et al.*, 1984; Holm *et al.*, 1984). Support of this notion can be derived from the retroviral aspartic proteinases which are dimeric, but have essentially the same active site as their cellular counterparts (Pearl & Taylor, 1987; Miller *et al.*, 1989; Toh *et al.*, 1985). The monomers of the viral enzymes are about half the size of eukaryotic aspartic proteinases.

All non-retroviral aspartic proteinases are synthesized as inactive zymogens that have a highly basic prosegment of 43 to 47 amino acids (Foltmann, 1988a). In the present paper, residues in the prosegment of human progastricsin (hPGC) are given a "p" suffix, and are numbered from 1p to 43p where 43p is the amino acid on the N-terminal side of the peptide that is cleaved during conversion. The amino acids corresponding to the mature enzyme are numbered from 1 through to the C-terminus. In (hPGC), the prosegment extends from Ala1p to Leu43p, and the mature enzyme portion from Ser1 to Ala329. Unlike most other families of proteinase zymogens which require activation by a second proteinase, the aspartic proteinase zymogens are self converting, the conversion to the mature enzyme being triggered by a drop in pH from neutrality to acid conditions (Herriot, 1939). The lowering of the pH protonates acidic residues on the mature enzyme portion of the molecule, thereby disrupting favorable electrostatic interactions with positively charged amino acids on the prosegment (Herriot, 1939;

---

[1]A version of part of this chapter has been submitted for publication. James *et al.*, 1994, *Structure and Function of the Aspartic Proteinases.* (In press).

Perlmann, 1963; James & Sielecki, 1986). Subsequent conformational changes in the prosegment of pepsinogen and in the first fourteen residues of the mature enzyme lead to one or more intramolecular proteolytic cleavages that liberate the prosegment from the zymogen (McPhie, 1972; Al-Janabi *et al.*, 1972; Bustin & Conway-Jacobs, 1971; Kageyama *et al.*, 1989; Foltmann & Jensen, 1982; Foltmann, 1988b). Above pH values of 4, the conversion process appears to be predominantly bimolecular, pepsin cleaving pepsinogen. Below pH 3, conversion is a unimolecular process.

The gastric juice of mammals contains two major aspartic proteinases, gastricsin and pepsin. Pepsin is by far the most abundant of the two and is produced by the chief cells of the oxyntic glands (Foltmann, 1981). These glands are found throughout the major portion of the stomach (the fundus) and are also responsible for acid and mucus secretion. Gastricsin is not only secreted by the chief cells of the fundus, but it can be found associated with the pyloric glands of the antral mucosa (Foltmann, 1981). Pepsinogens are generally more slowly activated than progastricsins and pepsins generally carry more negatively charged residues, suggesting that they have a lower pH optimum than gastricsins (Bank *et al.*, 1991). Both zymogens are activated upon exposure to the acidic medium of the stomach (Foltmann, 1981; Samloff, 1982). hPGC and human pepsinogen (hPGN) are found in the circulation of humans but only pepsinogen is present in human urine, as circulating progastricsin is readsorbed on the glomerular membrane of the kidney (ten Kate *et al.*, 1989). Gastricsins and pepsins generally prefer substrate peptides that contain hydrophobic amino acids at the P1 and P1' positions (nomenclature according to Schechter & Berger, 1967). In general, aspartic proteinases display kinetics such that k_{cat} and not k_m depends on the length of the substrate (Fruton, 1976). Only peptides of four amino acids or longer are effectively hydrolyzed by most aspartic proteinases. Gastricsin and pepsin follow this generally observed trend.

The gene for human progastricsin is encoded at a single locus on chromosome 6 (Taggart *et al.*, 1989; Pals *et al.*, 1989), whereas the human pepsinogen genetic locus is polymorphic and codes for at least three distinct polypeptide sequences on chromosome 11 (Taggart *et al.*, 1985). The genes encoding hPGC and hPGN have been cloned and sequenced (Hayano *et al.*, 1988; Sogawa *et al.*, 1983). The predicted polypeptide products are clearly homologues, displaying 45% amino acid identity when aligned (Hayano *et al.*, 1988).

The three dimensional structures of porcine pepsinogen (James & Sielecki, 1986; Sielecki *et al.*, 1991) and porcine pepsin (Sielecki *et al.*, 1990) have been solved in this laboratory at high resolution using X-ray diffraction techniques. Other structure determinations of porcine pepsinogen (Hartsuck *et al.*, 1992) and of porcine pepsin (Abad-Zapatero *et al.*, 1990; Cooper *et al.*, 1990) have been reported. The prosegment of pepsinigen is predominantly helical except for a β-strand at the N-terminus (James & Sielecki, 1986). Electrostatic, hydrogen bonding and hydrophobic interactions contribute to the binding between the prosegment and the rest of the protein (Sielecki *et al.*, 1991). The majority of these interactions with the prosegment are made by amino acids 7 to 12 of the mature enzyme. Interestingly, this five residue segment is found to occupy the active site in the zymogen; it moves up to 40 Å to become part of a six-stranded antiparallel β-sheet upon conversion to the mature enzyme. This appears to be a conserved feature of aspartic proteinases. In an effort to understand the similarities and differences between progastricsins and pepsinogens more fully, and to study further the process of activation in aspartic proteinase zymogens, we have undertaken to solve the three-dimensional structure of human progastricsin (EC 3.4.23.3) using X-ray crystallography.

MATERIALS AND METHODS

Crystallization

The isolation, purification and preliminary crystallization of progastricsin have been described (Ivanov *et al.*, 1990). The crystals are tetragonal and belong to the space group $P4_22_12$. We were unsuccessful in reproducing the original crystals from a fresh protein preparation and resorted to using crystal seeds from a previous preparation. Seed crystals from the original phosphate mother liquor were used to grow fresh crystals from a sodium formate mother liquor, pH 6-6.5. Seeded 10 µL drops of protein at 50 mg/mL in 0.63 M sodium formate and 20 mM Na/K phosphate were equilibrated against 1.8 M sodium formate with 3-4% v/v ethanol. Crystals grew to dimensions of 0.5 x 0.5 x 0.5 mm^3 within 3 weeks and diffracted to 1.6 Å resolution. The space group and unit cell dimensions (Table6.1a) were indistinguishable from those previously published (Ivanov *et al.*, 1990).

Table 6.1a

Crystallographic data for human progastricsin

Derivative Soak Conditions‡		Native	Ethyl Mercury Phosphate 10 mM, 1 day	Uranyl Acetate 10 mM, 2 days	Mersalyl 10 mM, 2 days	Trimethyl Lead Acetate 50 mM, 2 days
Unit Cell Dimensions (Å) ($P4_22_12$)	*a=b*	105.31	105.10	105.09	105.32	105.30
	c	70.42	70.33	70.60	71.01	70.25
Number of unique reflections measured (15-3.0 Å)		7829	7553	7667	7542	7767
Anomalous differences			6425	6572	0	6780
Average redundancy of measured intensities		9.3	10.6	11.8	13.6	11.4
Percent of theoretically observable reflections		93.7%	90.4%	91.8%	90.3%	93.0%
Decay estimate		18%	11%	3%	16%	Not measured
R_{merge}*		3.1%	6.2%	5.9%	9.1%	2.9%
Isomorphous (R_{iso}) differences with native data§			16.2%	16.7%	20.1%	16.3%
Phasing power+			1.56	0.67	1.17	1.59
FOM (15-3 Å)Δ		0.76 (0.63 for enantiomorph)				

Table 6.1a

‡ The soaking buffer was 1.8 M sodium formate, pH 6.0, for all heavy atom derivatives

* $$R_{merge} = \frac{\sum_{hkl}\sum_{i} |I_{i,hkl} - \langle I \rangle_{hkl}|}{\sum_{hkl}\sum_{i} I_{i,hkl}}$$

§ $$R_{iso} = \frac{\sum_{hkl} ||F_P| - |F_{PH}||}{\sum_{hkl} |F_P|}$$, where $|F_P|$ and $|F_{PH}|$ are the structure factor amplitudes for the native and heavy-atom derivatised crystals.

\+ $$\text{Phasing Power} = \left(\frac{\sum_{hkl} |F_H|^2}{\sum_{hkl} (|F_{PH}(obs)| - |F_{PH}(calc)|)^2} \right)^{1/2}$$

Δ FOM: Figure of Merit.

Data collection

X-ray diffraction data for one native data set to 1.62 Å resolution and four heavy atom derivatives to 3.0 Å resolution were collected on a twin San Diego multiwire detection system at the University of Alberta (Xuong *et al.*, 1985). Graphite monochromated $CuK_{\bar{\alpha}}$ X-rays were produced by a Rigaku RU-200 rotating anode X-ray generator operating at 6 kW power (40 kV, 150 mA). The collimated X-ray beam was 0.7 mm in diameter. Each data set was collected from one crystal over a period of about 24 hours. All data sets were processed with the San Diego software (Howard *et al.*, 1985). For the heavy atom derivative data, excluding mersalyl, Bijvoet pairs were not merged in any of the data processing steps. The native data set was put on an absolute scale using the computer program ORESTES (Thiessen & Levy, 1973) and derivatives were scaled to the native data using a non-linear least squares procedure within the BIOMOL suite of programs.

Molecular replacement

The molecular structure of porcine pepsinogen at 1.8 Å resolution, provided the template for building a model of human progastricsin (Bank *et al*, 1991) using the human progastricsin gene sequence (Hayano *et al.*, 1988). A fast rotation function (Navaza, 1994) was used to search Patterson space for the orientation of the model. A consistent peak was found with both the human progastricsin model, and the refined porcine pepsinogen coordinates at Euler angles $\alpha = 19.2°$, $\beta = 56.2°$, $\gamma = 294.5°$. A typical solution using integration radii from 4-28 Å and data with $|F_o| > 3\sigma |F_o|$ between 10 and 3.0 Å resolution was 5.3 standard deviations above the mean and 1.0 standard deviation above the next highest peak.

The rotated model was then used to search translation space using the computer program Brute (Fujinaga & Read, 1987). However, no significant peaks were found in the translation function, and we proceeded to collect heavy atom derivatives to assist in the phasing process.

Heavy atom derivatives

Four potential heavy atom derivatives having reasonable values for their corresponding isomorphous differences to the native data (Table 6.1a) were found for progastricsin. The derivatives included ethyl mercury phosphate (EMP), mersalyl,

Table 6.1b

Heavy atom sites for Human Progastricsin

	Site	x	y	z	occ*	anocc*	B_{iso} (Å^2)†
Ethyl mercury phosphate	1	0.3272	0.1706	0.2188	58.1	4.5	31.7
	2	0.2664	0.3277	0.3025	56.0	4.0	65.5
	3	0.3827	0.3467	0.4305	34.0	2.9	67.0
	4	0.3197	0.3257	0.1643	30.7	2.3	5.88
Uranyl acetate	1	0.0516	0.2146	0.1347	67.4	6.6	54.9
Mersalyl	1	0.3189	0.3253	0.1650	66.6	0.0	9.66
Trimethyl lead acetate	1	0.1091	0.4262	0.1677	50.8	4.4	21.1
	2	0.1290	0.2684	0.3488	49.8	4.3	39.9
	3	0.2769	0.3304	0.2759	24.7	2.1	45.4
	4	0.1717	0.4160	0.3398	52.5	5.2	86.7
	5	0.3469	0.0275	0.2184	21.7	1.7	36.5

* Occupancies and anomalous occupancies are given in electrons.

† $B_{iso} = 8\pi^2/3 \langle r^2 \rangle$

uranyl acetate, and trimethyl lead acetate. Inspection of difference Patterson maps yielded a single site for the uranyl derivative and one of the ethyl mercury phosphate sites also. The resulting heavy atom parameters were refined with MLPHARE (Otwinowski, 1990) and the two sites were put on a common origin using cross Fouriers. Cross Fouriers yielded another EMP site and four sites for the lead derivative. Subsequent cycles of heavy atom refinement, checking with cross Fouriers and anomalous Fouriers yielded eleven sites (Table 6.1b) for the four derivatives giving a figure of merit of 0.76 for all data from 15-3 Å resolution for the correct heavy atom enantiomorph. The incorrect enantiomorph yielded a figure of merit of 0.63. The heavy atom phases from MLPHARE were used to calculate a phased translation function (Read & Schierbeek, 1988) for the rotated progastricsin model. The result was a very clear translation vector peak at x = 0.8083, y = 0.6886, z = 0.7075. The peak height was 30 σ above the mean value for the phased translation map. Placing the rotated and translated model in the MIR map displayed clearly the main features of the molecule including the hydrophobic core and Lys37p coordinated by the two active site aspartates. However, even with this unusually good signal and relatively high figure of merit, it was evident that many of the surface loops were disordered and not visible in the MIR map.

After one round of molecular dynamics refinement with XPLOR (Brünger *et al.*, 1987), the R-factor dropped from 0.55 to 0.30 for all data with $|F| \geq 3\sigma |F|$ between 10 and 3.0 Å resolution (7829 observations). After two rounds of model building with FRODO (Jones, 1978), phase combination of the refined model and MIR phases with SIGMAA (Read, 1986) and XPLOR refinement, the R-factor dropped to 0.28 for all data with $|F| \geq 3\sigma |F|$ between 10 and 2.0 Å (25306 observations). Further refinement with PROLSQ (Hendrickson & Konnert, 1980) and TNT (Tronrud *et al.*, 1987; Tronrud, 1992), using all data between 20 and 1.62 Å, reduced the R-factor to 0.179. The final refined molecule includes all 372 amino acid residues of the zymogen, and 250 solvent molecules. Refinement statistics are presented in Table 6.2. A Ramachandran plot of the refined model is shown in Figure 6.1. All residues are found in stereochemically allowed regions of the (ϕ,ψ) map.

Analysis of the final model with PROCHECK (Laskowski *et al.*, 1993) puts our structure of hPGC within acceptable ranges of stereochemical criteria for a protein structure determined at 1.62 Å resolution. Coordinates of the refined hPGC molecule and solvent molecules have been deposited with the Brookhaven Protein Data Bank (Bernstein *et al.*, 1977) under the accession code 1hpc. Figures 6.2 and 6.3 show two

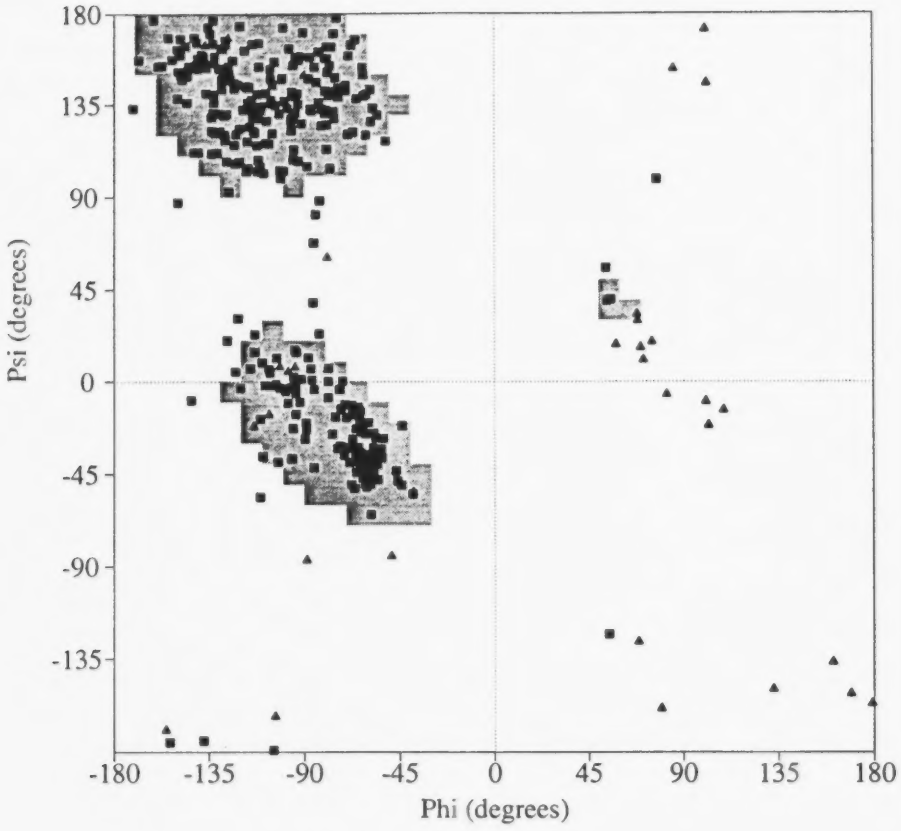


Figure 6.1 A Ramachandran plot of the refined human progastricsin model is shown. The most favorable regions of the map are shaded. Glycine residues are marked with triangles and all other residues are shown as squares (Laskowski *et al.*, 1993).

Table 6.2

Crystallographic refinement statistics for human progastricsin

Resolution range (Å)	20.0 - 1.62
R factor*	0.179
Number of reflections	50353
Completeness of data	96%
‡R_{merge}	5.9%
Total no. of protein atoms	2863
Total no. of solvent molecules	250
rms deviations from ideal stereochemistry:	
bond distances	0.019 Å
bond angles	2.4°
planar groups	0.016 Å
ω of peptide bonds	3.6°

* R-factor = $\Sigma||F_O|-|F_C||/\Sigma|F_O|$ where $|F_O|$ and $|F_C|$ are observed and calculated structure factor amplitudes, respectively.

‡ See Table 6.1a for definition.

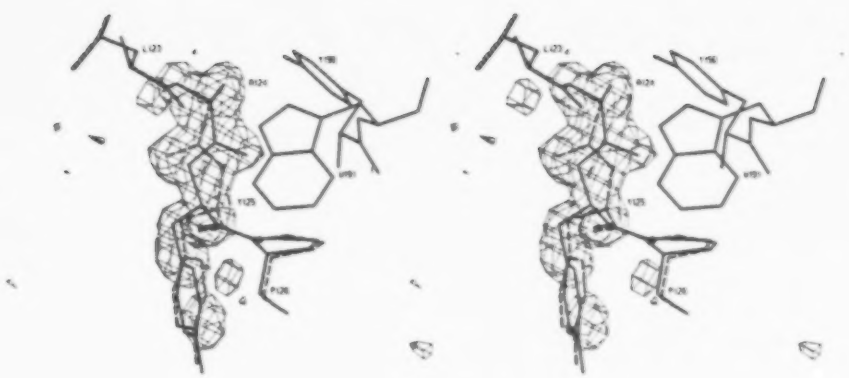


Figure 6.2. A $2|F_o|-|F_c|$ omit map of the region containing the peptide between residues Ala124 and Tyr 125. This peptide is part of a disordered loop in hPGC. The loop is in a different conformation in pPGN. The peptide has two distinct conformations. The refined model conformation is shown in solid thick lines and a possible second conformation is shown as the dashed line. Both conformations allow the side chain of Tyr125 to occupy the same position in space. The map is contoured at 0.23 e/Å^3 (σ = 0.07 e/Å^3).

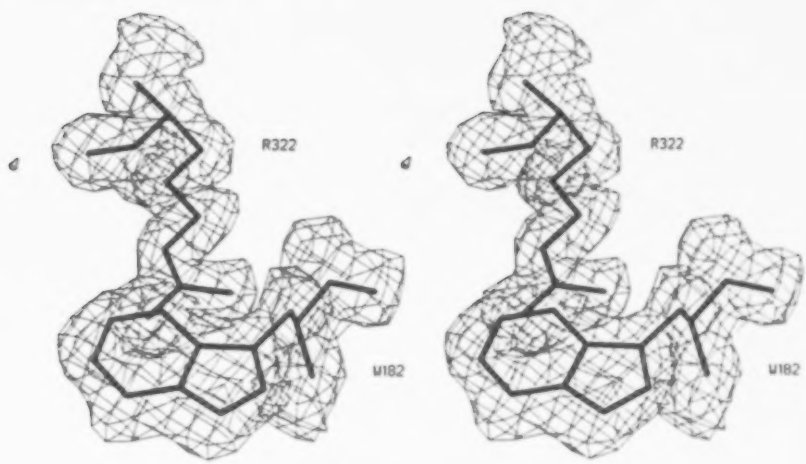


Figure 6.3. A portion of the $3|F_o|-2|F_c|$ electron density map around the refined model at amino acids Arg322 and Trp182. The map is contoured at 0.35 e/Å^3. Both residues are conserved in progastricsin amino acid sequences and are on the surface of the six-stranded antiparallel β-sheet.

Table 6.3

Disordered regions in the structure of human progastricsin

Residue Range	B-factors ($Å^2$)†	Comments
Glu25p-Lys31p	52-76	Breaks in main-chain density, poor side-chain density.
Arg39p-Leu43p	58-79	Breaks in main-chain density, poor side-chain density.
Tyr75-Ser79	47-65	Poor side chain density (tip of flap).
Ala127-Thr134	50-100	Breaks in main-chain density, poor side-chain density. Region displaced by pro-segment C-terminus.
Asn157-Ser161	69-83	Breaks in main-chain density, poor side-chain density.
Trp206-Glu210	47-54	Poor side chain density in region of S-S bridge.
Glu243-Tyr244	46-50	Poor side chain density.
Asn280-Gly282	53-63	Poor side chain density, small breaks in main chain.

† Average B-factor ranges for the main-chain atoms of the residues involved.

regions of the model and their associated electron density. Figure 6.2 illustrates part of a loop that has two alternate conformations of the main-chain peptide between amino acids Ala124 and Tyr125. Our refined model contains only one of the conformers where Tyr125 is just on the outside edge of allowed β conformations (ϕ=-54, ψ=114). Tyr125 is just outside the α_L region (ϕ=88, ψ=92) in the other conformation. Only the positions of the main chain atoms differs in the two models, the side chain being in nearly the same place. One would expect the β conformer of Tyr125 to be in a lower energy state by about 8 kilocalories, but both species appear to have equal electron density. There are no additional interactions that appear to stabilize the α_L peptide or to destabilize the β peptide at Tyr125.

Figure 6.3 displays two well ordered side chains, Trp182 and Arg322. Both residues are part of a conserved six-stranded antiparallel β-sheet found in all aspartic proteinases and their zymogens.

RESULTS

Overall description of the progastricsin molecule

Not surprisingly, the molecular structure of human progastricsin is very similar to that of porcine pepsinogen (the aligned protein amino acid sequences display a 45% identity). A C^{α} tracing of pPGN superposed onto hPGC is shown in Figure 6.4. Porcine pepsin is overlaid onto hPGC in Figure 6.5. The root mean square difference (rmsd) between 1176 equivalent main chain atoms (294 residues) for the two zymogens is 1.30Å. The corresponding value for 269 eqivalent residues between hPGC and porcine pepsin is 1.35Å. As in other aspartic proteinases, the structure of the N-terminal domain of the hPGC molecule (residues Ser1 to Leu150) is more structurally conserved than the C-terminal domain (residues Phe151 to Ala329). This is partly due to the C-terminal domains generally having greater thermal mobility. Although hPGC and pPGN have very similar overall structures, there are striking differences in several regions of the molecules. Regions where there are large rigid body motions of secondary structural units, or where totally different conformations are exhibited (greater than 3.0 standard deviations in C^{α} distances during the superposition), are illustrated in Figure 6.6. For hPGC, this cutoff distance is about 3 Å. Most of the regions having dissimilar conformations in the two zymogens are also partially disordered and have poorly defined electron density and high thermal motion factors. Disordered regions in hPGC are

Figure 6.4. A superposition of porcine pepsinogen onto human progastricsin. Only the C^{α} atoms for both models are shown. The equivalent amino acids are given in Table 6.4. The rmsd for 1176 main-chain atoms in 294 equivalent residues is 1.30 Å. The progastricsin molecule is shown as a solid thick line. The prosegment is drawn with a thicker line than that depicting the rest of the molecule. Every twentieth amino acid is numbered in hPGC. The pPGN model is shown in dashed lines.

Figure 6.5. A superposition of porcine pepsin onto human progastricsin. The rmsd for 1076 main-chain atoms in 269 equivalent residues is 1.35 Å. The models are drawn and labeled as in Figure 6.4.

Table 6.4

Equivalent polypeptide segments in human progastricsin and porcine pepsinogen

hPGC	pPGN
V3p - Y38p	V2p - Y37p
P6 - F71	L6 - L71
L80-A124	M80-A124
T135 - Q159	P135 - D159
G164 - Q203	S163 - E202
G211 - Q227	G209 - Thr225
N252 - I277	S250 - I275
V286 - L293	S284 - V291
Q299 - A329	G296 - A326

In hPGC, the polypeptide follows a drastically different path from that seen in pPGN in the amino acid segments 39-5p, 125-134, 160-163, 204-210 and 294-298. Rigid body movements are seen in the segments 71-79, 228-251, and 278-285. There are three one residue insertions in hPGC at positions Ser161, Gly206 and Ser294, relative to the porcine pepsinogen amino acid sequence and structure. The first two insertions are located at disulphide bridges and the third is on a surface loop. All three insertions contribute to the different conformations seen in hPGC relative to pPGN. Equivalence was determined by main chain atom pairs that differed by less than three standard deviations in the superpositions. The rmsd is 1.30Å for 1176 equivalent main chain atoms (294 residues).

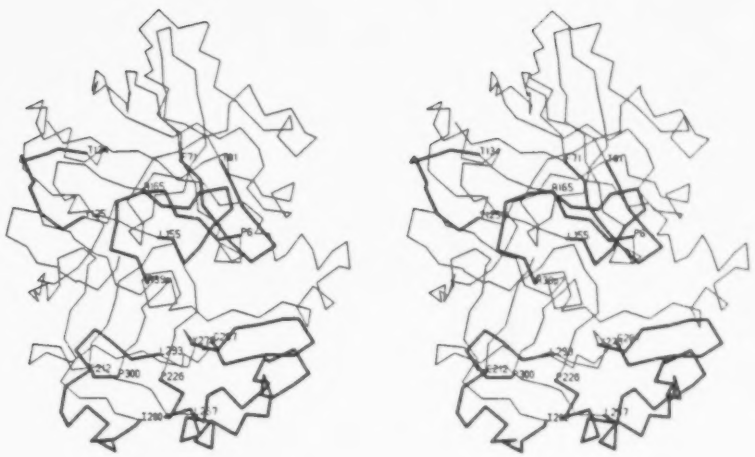
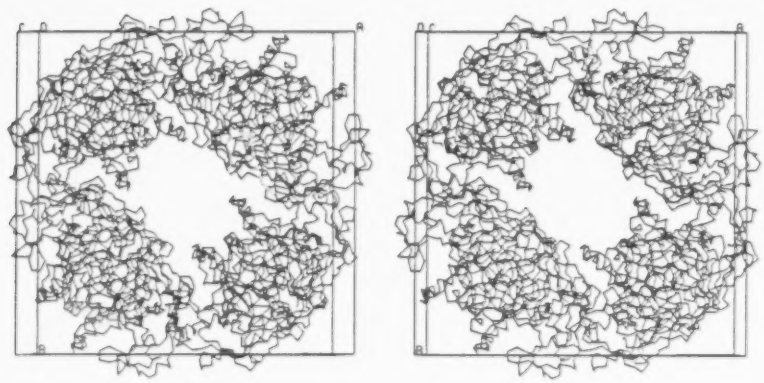
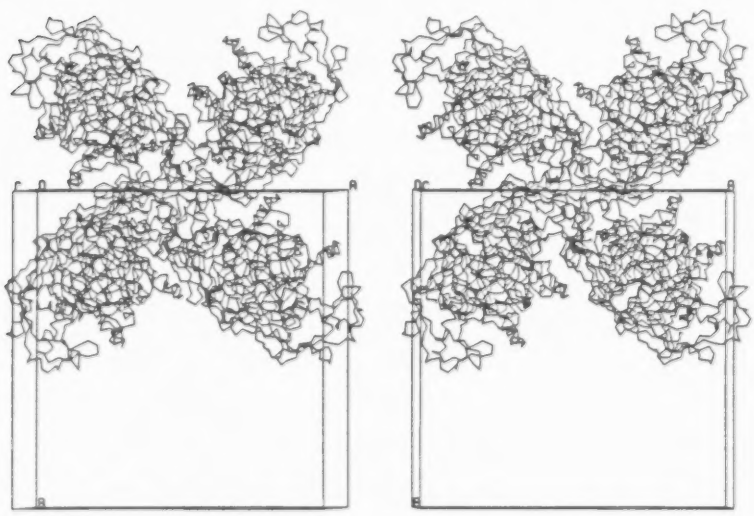


Figure 6.6. A C^{α} representation of the hPGC model highlighting (in thick lines) amino acid segments that are displaced relative to porcine pepsinogen. The loops containing Arg39p-Glu5, Tyr125-Thr134, and Ala204-Gln212 are in totally different conformations in the two zymogens. The other regions are related by rigid body movements and the main chains are in essentially the same local conformation.

Figure 6.7. Molecular packing in the tetragonal unit cell of hPGC. C^{α} traces of hPGC are shown along with the unit cell.



(a) The packing around the solvent channel is illustrated. There are eight hPGC molecules in the unit cell. The channel is a 33 x 38 Å ellipse. The view is looking down the crystallographic c axis.



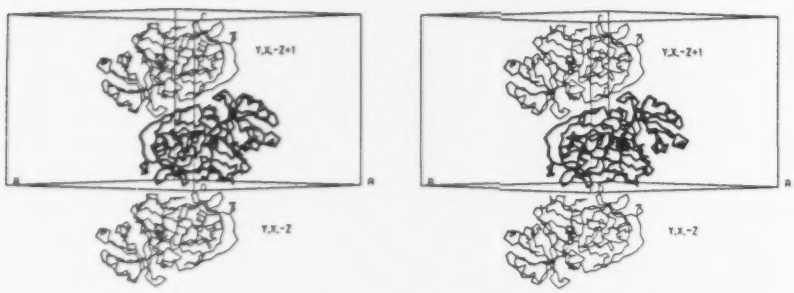
(b) The packing around the 4_2 screw axis at 1/2,0,z is illustrated. The view direction is similar to that shown in (a).

Table 6.5

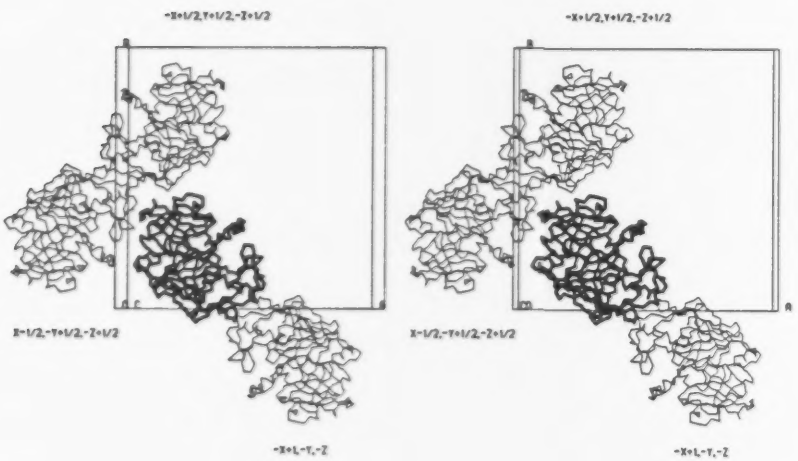
Protein-protein contacts in human progastricsin involving crystallographic symmetry

Molecule at x,y,z	Transformed Molecule at x',y',z'	Transformation Mapping x,y,z onto x',y',z'	Number of Protein-Protein Contacts ≤ 4 Å
D33p, W36p	S72, L73		
R39p	**F40p, G41p, D42p,** S1, V2, T3	y,x,1-z	65 contacts
Q69, T70	Y292, S294, Q227	(2-fold axis on x,y diagonal at x,x,1/2)	9 H-bonds
Q74, G76	Q74, G76		
Q92, S93	Q92, S93	y,x,-z	10 contacts
V3p	S93	(2-fold axis on x,y diagonal at x,x,0)	2 H-bonds
P23	S162		
G201, G202	A236, N256, L257	1-x,-y,-z	31 contacts
P258, S259	N256, P258	(2-fold axis coincident with 4_2 axis at 1/2,0,z)	3 H-bonds
Y64, S65	N250, G282, Y283	x-1/2,1/2-y,1/2-z	39 contacts
Y86, N98	G282, Y283	(2_1 screw axis parallel to x at x,1/4,1/4)	4 H-bonds
S62	N240		
T66	N252		
Y44	Y181, V267, E268	1/2-x,1/2+y,1/2-z	28 contacts
Q46	W182	(2_1 screw axis parallel to y at 1/4,y,1/4)	3 H-bonds
E105	Q179		

Figure 6.8. The five neighboring hPGC molecules and their symmetry relation to the parent molecule are shown (see Table 6.5). Only neighbors making direct protein-protein contacts are shown.



(a) Packing parallel to the crystallographic c axis.



(b) Packing in the ab plane.

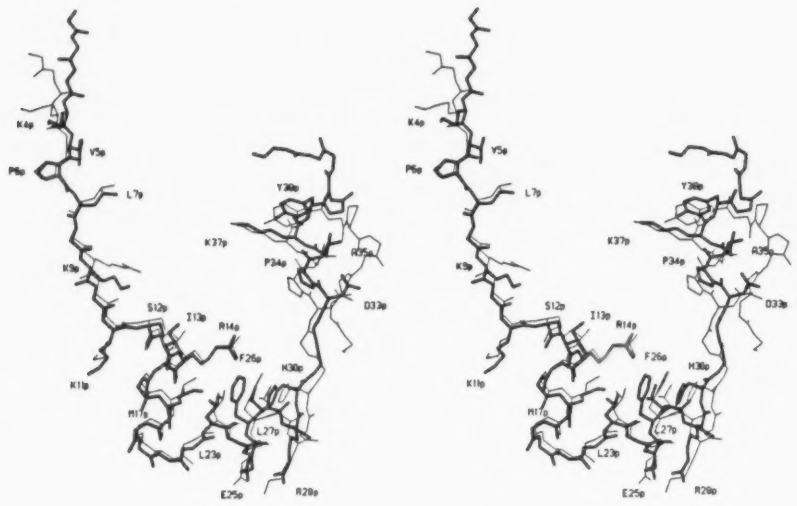
summarized in Table 6.3. Structurally equivalent polypeptide segments in hPGC and pPGN, along with the positions of three one-residue insertions in hPGC relative to pPGN, are given in Table 6.4. Several of the disordered regions in human progastricsin face into large solvent channels centred around the two-fold axes at (0,0,z) and (1/2,1/2,z) parallel to the crystallographic c axis. These channels are approximately 33 x 38 Å in cross section and run through the whole crystal (Figure 6.7).

Table 6.5 contains a detailed analysis of the hPGC molecular packing in the crystal lattice. There are five nearest neighbor molecules that make protein-protein van der Waals contacts ≤4 Å with the primitive molecule at (x,y,z). Two of the neighboring molecules pack parallel to the crystallographic c axis and are generated by two-fold axes along the body diagonal at 0 and 1/2 in c (Figure 6.8a). The other three neighboring molecules are packed in the x,y plane and are related by 2_1 axes parallel to the a and b axes respectively, and by the 4_2 screw axis (Figure 6.8b). The most extensive contact surface (see Table 6.5) contains two loops that differ in conformation between hPGC and pPGN. These are the peptides containing residues Phe40p-Glu5 and the β-hairpin containing Tyr75 (Table 6.4 and Figure 6.6). Some of the structural rearrangements in these two regions could thus be due to crystal packing forces. All but one of the molecular contacts are well ordered and associated with well-defined electron density. The exception is the contact involving the loops that differ in conformation between hPGC and pPGN. Electron density is poorly defined in several regions of this interface (Table 6.3).

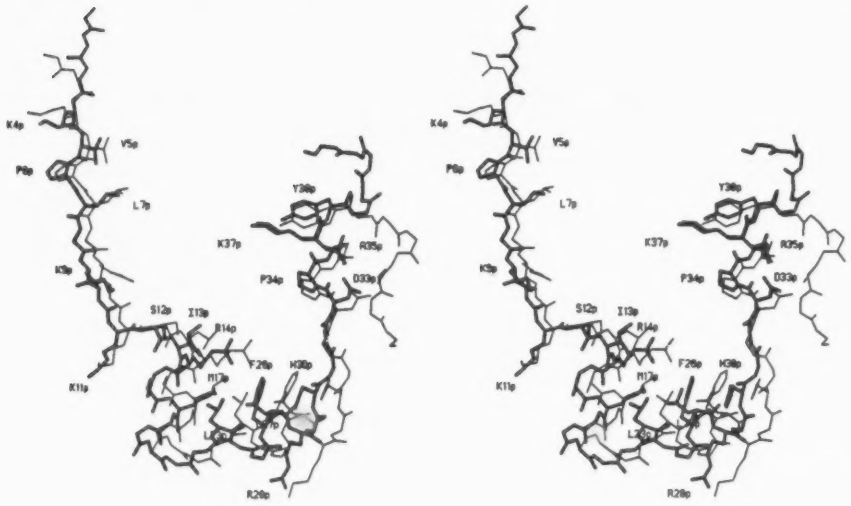
Analysis of the prosegment

The prosegment of hPGC consists of amino acids Ala1p-Leu43p, and is very similar in overall structure and conformation to the prosegment of pPGN. When superposed independently of the gastricsin and pepsin regions, the rmsd for 144 prosegment main-chain atoms between residues V2p-Y38p of hPGC is 1.1 Å. This superposition is shown in Figure 6.9a where the side chains for residues conserved in the prosegments of all known gastric proteinase zymogen amino-acid sequences (Figure 6.10) are displayed along with the main chain atoms of hPGC and pPGN. Optimising the superposition of the complete zymogens leads to a poorer overlap for the prosegments (rmsd= 1.5 Å for Val2p-Tyr38p) (Figure 6.9b). The main chain hydrogen bonding pattern, both within the prosegment and with β-strands in the rest of the enzyme, is nearly identical to that observed in pPGN (Table 6.6 and Figure 6.11).

Figure 6.9. A comparison of the prosegments of hPGC and pPGN.



(a) The prosegment of pPGN is overlaid onto the prosegment of hPGC. The rmsd for main chain atoms is 1.1 Å. Only side chains conserved in gastric proteinase prosegments are shown. hPGC is drawn in thick lines and pPGN is thin.

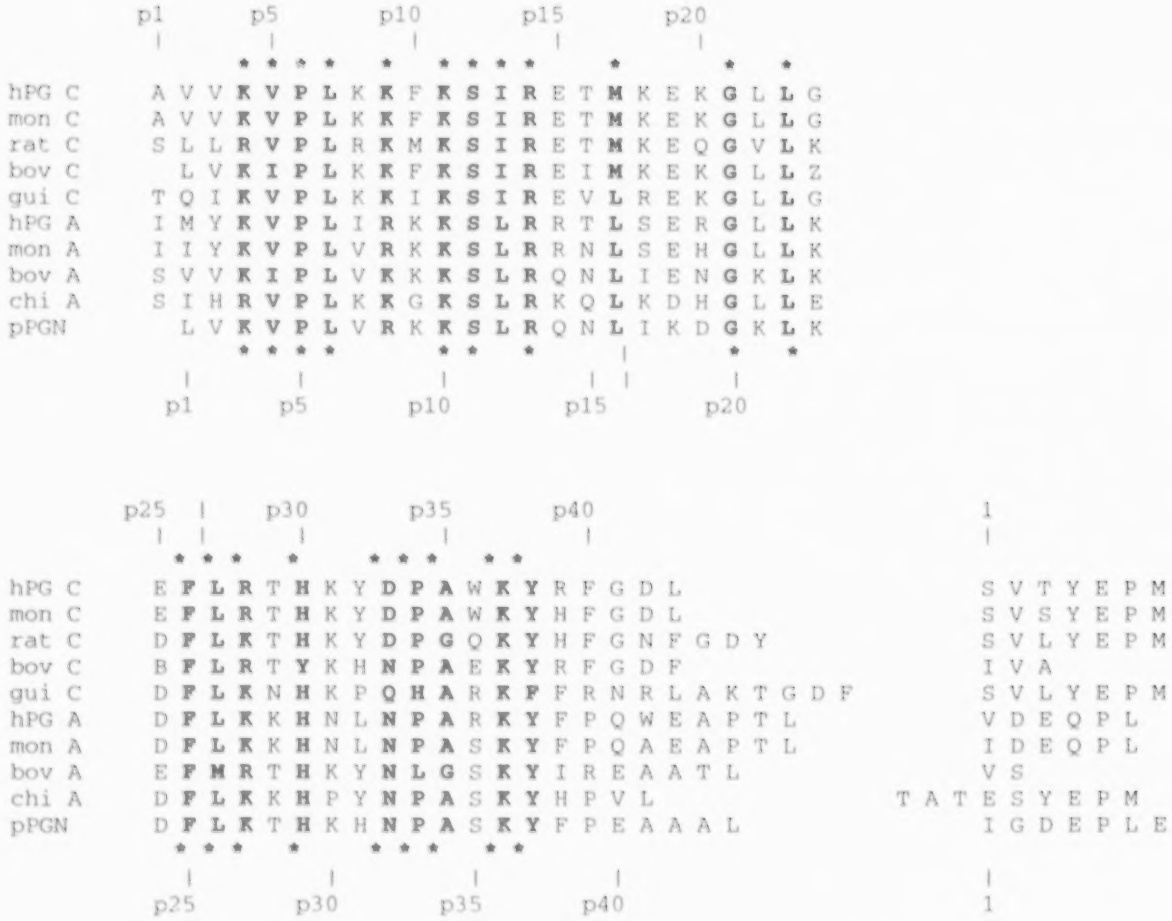


(b) The entire pPGN molecule is overlaid onto hPGC as per Figure 6.4. Only the prosegments are shown. The diagram is drawn as in Figure 6.9a. The rmsd for main chain atoms is 1.5 Å

Figure 6.10. Amino acid sequence alignment for the prosegments of pepsinogens and progastricsins (Foltmann, 1988a, 1988c). Conserved residues are highlighted in bold. The hPGC numbering is given at the top of the figure and that of pPGN is given at the bottom. The initial cleavage site in both prosegments is highlighted by a vertical bar adjacent to the position in the sequence. Sequences are: hPGC, human progastricsin (Hayano *et al.*, 1988; Taggart *et al.*, 1989); mon C, monkey progastricsin (Kageyama & Takahashi, 1986a); rat C, rat progastricsin (Ichihara *et al.*, 1986); bov G, bovine progastricsin (Foltmann, 1988a); gui C, guinea pig progastricsin (Kageyama *et al.*, 1992); hPG A, human pepsinogen (Sogawa *et al.*, 1983); mon A, monkey pepsinogen (Kageyama & Takahashi, 1986b); bov A, bovine pepsinogen (Harboe *et al.*, 1974); chi A, chicken pepsinogen (Baudys & Kostka, 1983); pPGN, porcine pepsinogen (Foltmann, 1988c and refs. therein).

Bovine progastricsin and bovine pepsinogen only have the amino acid sequences of their proparts determined. For the bovine progastricsin sequence, B denotes Asx and Z denotes Gln. Positions Trp42p and Glu43p of human pepsinogen are known to be variable (Foltmann, 1988c).

Figure 6.10


p1 p5 p10 p15 p20
hPG C A V V K V P L K K F K S I R E T M K E K G L L G
mon C A V V K V P L K K F K S I R E T M K E K G L L G
rat C S L L R V P L R K M K S I R E T M K E Q G V L K
bov C L V K I P L K K F K S I R E I M K E K G L L Z
gui C T Q I K V P L K K I K S I R E V L R E K G L L G
hPG A I M Y K V P L I R K K S L R R T L S E R G L L K
mon A I I Y K V P L V R K K S L R R N L S E H G L L K
bov A S V V K I P L V K K K S L R Q N L I E N G K L K
chi A S I H R V P L K K G K S L R K Q L K D H G L L E
pPGN L V K V P L V R K K S L R Q N L I K D G K L K
p1 p5 p10 p15 p20
p25 p30 p35 p40 1
hPG C E F L R T H K Y D P A W K Y R F G D L S V T Y E P M
mon C E F L R T H K Y D P A W K Y H F G D L S V S Y E P M
rat C D F L K T H K Y D P G Q K Y H F G N F G D Y S V L Y E P M
bov C B F L R T Y K H N P A E K Y R F G D F I V A
gui C D F L K N H K P Q H A R K F F R N R L A K T G D F S V L Y E P M
hPG A D F L K K H N L N P A R K Y F P Q W E A P T L V D E Q P L
mon A D F L K K H N L N P A S K Y F P Q A E A P T L I D E Q P L
bov A E F M R T H K Y N L G S K Y I R E A A T L V S
chi A D F L K K H P Y N P A S K Y H P V L T A T E S Y E P M
pPGN D F L K T H K H N P A S K Y F P E A A A L I G D E P L E
p25 p30 p35 p40 1

Two of the hydrogen bonds seen in hPGC (V3p,N to F168,O; K9p, N to S161,O) but not in pPGN are actually mediated via water molecules in the pPGN structure. There are also three conserved, highly ordered water molecules making identical hydrogen bonding interactions in both structures (Wat377, 380 and 419 of hPGC) (Figure 6.11). One of the few differences in the hydrogen bonding pattern arises because hPGC has an extra amino acid at the N-terminus. In hPGC the amino terminal nitrogen of Ala1p makes hydrogen bonds with the O^{γ} of Thr147 and the carbonyl oxygen of Ala145. Table 6.7 summarizes hydrogen bonds involving amino acid side chains in the prosegment of hPGC. Only about one half of the observed interactions are structurally conserved in porcine pepsinogen, although almost all of the amino acids participating in these hydrogen bonds are conserved in the sequences of the two proteins (Table 6.7). This observation would argue that secondary structural interactions are more strongly conserved features in proteins than tertiary interactions. This appears to be a general phenomenon (Chothia & Lesk, 1982).

Residues Ala1p to Lys11p in hPGC are in an extended conformation and make antiparallel β-sheet hydrogen bonds with two segments of the gastricsin moiety (Figure 6.11). Residues Val3p to Leu7p are hydrogen bonded to the segment Gly163 to Phe168 which is part of the second strand of the six-stranded antiparallel β-sheet that makes up the conserved core of all aspartic proteinases. Of course, Val3p to Leu7p makes up the first strand in the zymogen. In the active enzyme, the segment Val3p to Leu7p is likely replaced by amino acids Val2 to Tyr10 as is the case in porcine pepsin. Residues Lys4p to Leu7p are remarkably well conserved in all available amino acid sequences for the prosegments of pepsinogens and progastricsins (Figure 6.10).

Another antiparallel β-sheet is formed by residues Lys8p to Phe10p and Ala13 to Phe15. This interaction is noteworthy because residues 1 to 12 of hPGC and pPGN reside in the active site cleft in both zymogen structures and, in pPGN, undergo a remarkable conformational change upon acid conversion to pepsin. Residues 1 to 9 of pepsin move from the active site in the zymogen structure to the back of the molecule to replace the prosegment strand in the antiparallel β-sheet (James & Sielecki, 1986). In fact, most conserved interactions with the prosegments in hPGC and pPGN are made near the active site cleft with amino acids Met7 to Asp11. The most strongly conserved

Table 6.6

Hydrogen bonds involving only main chain donor and acceptor atoms in the prosegment of human progastricsin

Donor	Acceptor	Distance Å	Conserved H-Bond in pPGN
A1p,N	A145,O	3.2	NO
V3p,N	F168,O	2.8	NO
V5p,N	V166,O	2.9	YES
L7p,N	G164,O	2.9	YES
K9p,N	S161,O	2.5	NO
G163,N	L7p,O	2.7	NO
V166,N	V5p,O	2.9	YES
F168,N	V3p,O	2.7	YES
K8p,N	F15,O	3.0	YES
F10p,N	A13,O	2.8	YES
F15,N	K8p,O	3.0	YES
R14p,N	M10,O	2.9	YES
T16p,N	S12p,O	3.0	YES
M17p,N	I13p,O	2.9	YES
K18p,N	R14p,O	3.0	YES
E19p,N	E15p,O	3.0	YES
K20p,N	T16p,O	3.0	YES
G21p,N	K18p,O	2.8	YES
L22p,N	M17p,O	2.7	YES
F26p,N	L22p,O	3.3	YES
L27p,N	L23p,O	2.6	YES
R28p,N	G24p,O	3.0	YES
T29p,N	F26p,O	3.3	NO
W36p,N	D33p,O	2.8	YES
K37p,N	P34p,O	3.3	YES
Y38p,N	A35p,O	3.2	YES
R39p,N	W36p,O	3.5	NO

Figure 6.11. Hydrogen bonding interactions in the prosegment of hPGC (Tables 6.6 and 6.7) are illustrated. Regions of the mature enzyme that interact with the prosegment, mostly through antiparallel β-sheet hydrogen bonding, are included. The polypeptide chain is drawn as a thin solid line and the side chains participating in hydrogen bonds are shown as thick lines. The dashed lines are the hydrogen bonds. Three water molecule that make identical hydrogen bonds in the prosegments of hPGC and pPGN are also shown.

Table 6.7

Hydrogen bonds involving amino acid side chains in the prosegment of human progastricsin

Donor	Acceptor	Distance (Å)	Conserved H-Bond in pPGN	Conserved Amino Acids in pPGN
A1p,N	T147,O^{γ}	2.9	NO	NO
K4p,N^{ζ}	D172,$O^{\delta 2}$	2.6	NO	YES
K9p,N^{ζ}	Y14,OH	2.9	NO	YES
K9p,N^{ζ}	Q159,$O^{\varepsilon 1}$	3.0	NO	NO
K11p,N^{ζ}	E19p,$O^{\varepsilon 2}$	3.1	NO	YES
S12p,O^{γ}	E15p,$O^{\varepsilon 2}$	3.4	NO	YES
R14p,N^{ε}	D11,$O^{\delta 1}$	2.5	YES	YES
R14p,NH_2	D11,$O^{\delta 2}$	2.9	YES	YES
E15p,N	S12p,O^{γ}	2.9	YES	YES
T291,$O^{\gamma 1}$	D33p,$O^{\delta 1}$	2.7	NO	NO
Y292,N	D33p,$O^{\delta 2}$	2.9	NO	NO
K37p,N^{ζ}	D32,$O^{\delta 1}$	2.7	YES	YES
K37p,N^{ζ}	D217,$O^{\delta 2}$	2.8	YES	YES
K37p,N^{ζ}	O419,O	2.6	YES	YES
Y38p,OH	D217,$O^{\delta 2}$	2.5	YES	YES
T29p,$O^{\gamma 1}$	E25p,O	2.9	NO	NO
A35p,N	D33p,$O^{\delta 1}$	2.9	YES	YES
H30p,$N^{\delta 1}$	F26p,O	2.8	YES	YES

residues in this stretch are Tyr9 and Asp11. Tyr9 occupies the S1 pocket and makes a hydrogen bond with Asp32 at the active site, via its phenolic hydroxyl group in both zymogens. Tyr38p of the prosegment occupies the gastricsin S1' binding site and makes a hydrogen bond to Asp217 that is two-fold symmetrical to the Tyr9-Asp32 hydrogen bond (Figure 6.11). These two residues are thus related by the molecular pseudo two-fold axis that runs through the active site of all cellular aspartic proteinases (see Figure 6.12). Asp11 also makes a strong and conserved salt bridge with Arg14p (Figures 6.11 and 6.13). Even though there are seven conserved positively charged residues in the prosegments of all pepsinogen and progastricsin molecules sequenced to date, there are only two conserved salt bridges with the portions of the molecules that make up the mature enzyme. The Arg14p-Asp11 interaction is one of these, and the well known inhibitory double salt bridge between Lys37p and the catalytic residues Asp32 and Asp217 is the other. Most positively charged residues in the prosegment have either lysine or arginine at any one particular position. Lys11p, Arg14p and Lys37p however, are conserved more strongly than other charged sequences (Figure 6.10). Lys11p makes a good salt bridge with Glu19p in hPGC. This interaction possibly stabilizes the first helix in the prosegment. A similar interaction is possible in all of the other pepsin/gastricsin zymogens, except for porcine pepsinogen where the equivalent residue to 19p is a lysine.

There is an additional α-helix and one 3_{10} helix in the prosegment, all remarkably well conserved in the hPGC and pPGN structures. The first helix runs from Ile13p to Lys20p in hPGC and is capped at its N-terminus by the O^{γ} atom of the highly conserved Ser12p. This interaction appears to be a ubiquitous feature of aspartic proteinase zymogens. Gly21p is also a conserved residue and makes a classic C-capping interaction (Richardson & Richardson, 1988). In hPGC Glu15p makes a helix box capping interaction (Harper & Rose, 1993), but this feature appears only to be present in the progastricsin family (Figure 6.10), as a similar interaction is not found in pPGN.

The second α-helix spans residues Leu23p to Leu27p and runs roughly perpendicular to the direction of the prosegment's first α-helix (Figure 6.11). This second helix makes no hydrogen bonding interactions with the gastricsin segment of the molecule (Tables 6.6 and 6.7) and has high thermal-motion factors and weak electron density (Table 6.3). The main-chain nitrogen atom of Thr29p caps the penultimate carbonyl group of the helix (Phe26p). This interaction is not observed in porcine

pepsinogen. However, another C-cap hydrogen bond between the $N^{\delta 1}$ atom of His30p and the carbonyl oxygen of Phe26p is conserved in the two zymogen structures (Table 6.7 and Figure 6.11). Nonetheless, it should be noted that the electron density for these two residues in both structures is of poor quality (Table 6.3), indicating these residues are disordered in the crystal. The polypeptide chain of the prosegment exhibits high thermal motion parameters and discontinuous electron density from Glu25p to Lys31p in the hPGC structure. There are no hydrogen bonding interactions with the gastricsin portion of hPGC and the polypeptide following the second α-helix appears to be in an extended conformation.

Lastly and most importantly, a 3_{10} helix runs from Pro34p to Arg39p and contains two residues critical for inhibition of the zymogen: Lys37p and Tyr38p. The $O^{\delta 1}$ atom of Asp33p makes an N-cap hydrogen bond (d=2.8 Å) to the amide nitrogen of Ala35p. This hydrogen bond can also be classified as a type III Asx turn (Rees *et al.*, 1983). An identical hydrogen bond is made by Asn32p in pPGN, and appears to be a conserved structural feature in all but one of the gastric aspartic proteinase zymogens (Figure 6.10). Pro34p is the first residue of the 3_{10} helix and is strongly conserved,, along with Ala35p, Lys37p and Tyr38p. In this segment, only Trp36p is poorly conserved in the zymogen sequences. 3_{10} helices are relatively infrequent in proteins, and one with four consecutive hydrogens bonds is exceptional, especially in a nearly all-β protein such as hPGC. However, a 3_{10} helix in this area is indispensable, as it is narrower than an α-helix, and can therefore easily fit into the active site groove. The 3_{10} conformation also facilitates the interaction of two adjacent amino acid side chains (Lys37p and Tyr38p) with the active site carboxylate of Asp217.

At Arg39p the polypeptide chain in hPGC follows a drastically different path from that observed in pPGN (Table 6.4 and Figures 6.4, 6.6 and 6.9). This marked conformational difference follows through until residue Pro6 in hPGC, where the two polypeptides again adopt a common path in the respective zymogens. It is interesting that Pro6 in hPGC aligns structurally with Leu6 of pPGN and not Pro5 (Figure 6.10). This rearrangement in hPGC relative to pPGN causes the active site flap containing Tyr75 to adopt a markedly different conformation than that exhibited in most other aspartic proteinases. Normally Tyr75 makes a hydrogen bond with the $N^{\varepsilon 1}$ atom of Trp39. The conformational change observed in the Arg39p to Pro6 segment appears to have a *domino* effect as the nearby loop Ala127 to Thr134 must move from its observed

Table 6.8

van der Waals contacts involving hydrophobic amino acid residues in the prosegment and mature enzyme regions of human progastricsin

Amino acid in prosegment	Amino acid in mature enzyme
V5p, L7p	Y14, G16, I18, V91, L155 G163, V166, F168
I13p	M10, A115, F117
R14p	M10
M17p, L23p, L27p	M10, Y114
F26p, H30p, Y32p	M7
Y38p	Y190, I215, I303

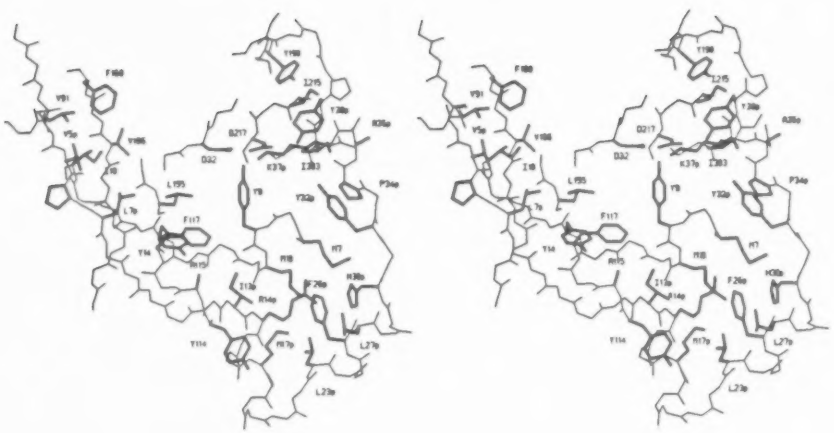


Figure 6.12. Amino acids contributing to hydrophobic interactions between the prosegment and mature enzyme segments of the hPGC molecule are shown. The main chain is drawn in thin and the side chains in thick.

position in pPGN to accommodate the different position of that segment. The 127 to 134 loop is disordered and has weak electron density (Table 6.3). Other regions of hPGC must also move to accommodate the repositioned active site flap. The surface helices in loops Gln43 to His53 and Thr110 to Gln116 make concerted movements away from the active site, relative to pPGN, to accommodate the displaced flap.

Aspartic proteinase zymogens contain another 3_{10} helix in the active site cleft, but this helix is at the N-terminus of the mature enzyme segment, spanning residues Glu5 to Met10. There is also a 3_{10} hydrogen bond between Ala12,N and Tyr9,O. The 3_{10} hydrogen bonding is interrupted between Ala8,O and D11,N, but a water molecule makes compensating hydrogen bonds. This 3_{10} helix sits in the active site cleft and makes a number of conserved interactions with the prosegment (Figures 6.11 and 6.14). Upon conversion to the mature enzyme this helix moves out of the active site and becomes part of the N-terminal β-strand of pepsin.

Along with hydrogen bonding and electrostatic interactions, there are a number of conserved hydrophobic contacts in the prosegment regions of hPGC and pPGN. These are summarized in Table 6.8 and Figure 6.14. Both the N-terminal β-strand and the first two α-helices are stabilized by hydrophobic interactions with the gastricsin portion of the zymogen. All amino acid residues listed in Table 6.8 are remarkably well conserved in the pepsinogen and progastricsin families of aspartic proteinase zymogens (Foltmann, 1988c). Note also that Tyr38p sits right in the highly hydrophobic S1' substrate binding pocket (Figure 6.15) that has been mapped in porcine pepsin (Sielecki *et al.*, 1991; Abad-Zapatero *et al.*, 1991; Chen *et al.*, 1992). Tyr9 makes van der Waals contact with the side chain of Ala115 and Phe117 in the S1 site (Figure 6.15). The methylene groups on Arg14p make van der Waals contacts with the side chain of Met10 (Figure 6.12). Hence Arg14p has a dual role in the stabilization of the prosegment, making both electrostatic and hydrophobic interactions with the mature enzyme moiety. Again many of the hydrophobic interactions come from the peptide segment Ser1 to Tyr14 of the mature enzyme that sits in the active site cleft in the neutral zymogen structure, and rearranges upon acidic activation to become a totally different structural part of the active enzyme.

Figure 6.13. The four conserved salt bridges are shown in hPGC. Side chains are illustrated in thick and the C^{α} trace in thin lines. The three disulphide bridges in hPGC are also shown. The prosegment is also drawn as a thick solid line.

Table 6.9

A. *Charged residues in the prosegments of human progastricsin and porcine pepsinogen*

Molecule	+ Charges	- Charges	Salt Bridges	Conserved Salt Bridges
hPGC	13	4	4	2
pPGN	14	3	7	2

On average both groups of zymogens have 13 positive charges in the propart segment, including the amino terminus (see Figure 6.10). The pepsinogen A family has, on average 3 negatively charged residues at neutrality, in comparison to an average of 4.5 negative charges for pepsinogen C's. There are two conserved salt bridges in the zymogen segments of each protein. For hPGC these are: Arg14p to Asp11, and Lys37p to Asp32 and Asp217 at the active site.

B. *Charged residues in the mature enzyme segments*

Molecule	Aspartic	Glutamic	Lysine	Arginine	Histidine
hPGC	7	16	0	3	1
pPGN	29	12	1	1	1

The mature enzyme segments each have in common two conserved salt bridges. For hPGC, these are: His51 to Asp118, and Arg310 to Asp306. In hPGC, and the other gastricsin zymogens, there is another conserved positively charged residue, Arg51. It is not conserved in the pepsinogen family of enzymes.

Electrostatic interactions in human progastricsin

In general, the pepsinogen A family (pepsins) contains more negatively charged amino acids than the pepsinogen C family (progastricsins) (Table 6.9). Also, pepsinogens have more aspartate than glutamate residues. The progastricsins generally have more glutamates than aspartates. However there seem to be only four conserved salt bridges in these two families of proteins (Figure 6.13). Two, already mentioned, are between the prosegment and the mature enzyme moieties, whereas the other two salt bridges are found within the mature enzyme segments of hPGC and pPGN (Table 6.9 and Figure 6.13). From the fact that the pepsin group is more negatively charged, one would expect it to exhibit a different pH activation profile than the gastricsin group, and also a different pH optimum. This could have something to do with the distributions of the two proteins in the stomach, as has been previously suggested (Foltmann, 1981). Several glutamate residues that are conserved in both the pepsinogen and progastricsin families don't have an apparent function in the hPGC structure. A conserved aspartate at position 87 in hPGC is buried from solvent and makes identical hydrogen bonds to other completely conserved amino acid residues in hPGC as in pPGN. This interaction has been noted previously in the structure of porcine pepsin (Andreeva & James, 1991). Asp306 can be found on a short α-helix near the active site and makes a conserved salt bridge with Arg310 one turn along on the same helix (Figure 6.13 and Table 6.9). Asp 118 is also conserved in all gastric proteinase sequences and makes a conserved salt bridge with His53 in both hPGC and pPGN. These conserved interactions have also been noted in an analysis of porcine pepsinogen (Hartsuck *et al.*, 1992).

The active site of human progastricsin

As already mentioned, the mechanism of inhibition for human progastricsin is entirely analogous to what is seen in porcine pepsinogen. The side chain ε-amino group of Lys37p makes electrostatic interactions with Asp32 and Asp217, displacing the nucleophilic water molecule found at the active site in mature aspartic proteinases. In addition Tyr38p is hydrogen bonded to Asp217 and occupies the S1' substrate binding site. In an interaction related by the active site pseudo-dyad axis, Tyr9 makes a hydrogen bond to Asp32 and sits in the S1' binding site. The $O^{\gamma 1}$ atom of Thr220 points away from Asp217 in an identical manner to what is observed in pPGN (Figure 6.15) (Sielecki *et al.*, 1991). There are important differences, however, at the active site

Figure 6.14. Separate stereoviews of hPGC (a) and pPGN (b) illustrate conformational differences in the active site region. Important amino acids at the active site are shown for both molecules, and the residues in hPGC are labeled. Residues Pro39p, Leu44, Ser127, and Gly132 on two disordered loops in pPGN are labeled for clarity. The analogous loops in hPGC are in drastically different conformations. This results in Tyr75 of hPGC not being accessible to the active site, where it resides in pPGN and most mature aspartic proteinases.

(a) hPGC

(b) pPGN

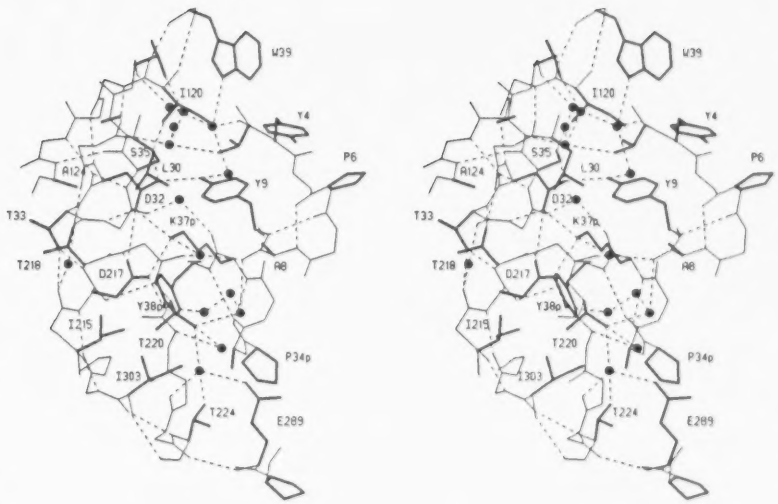


Figure 6.15. Hydrogen bonding interactions at the active site of hPGC. The figure is drawn as in Figure 6.11. Note that the segment Tyr4 to Pro6 prevents Tyr75 of the active site flap from sitting in the active site and making a hydrogen bond with the $N^{\varepsilon 1}$ atom of Trp39.

The peptide from Ser1 to Pro6 in hPGC is in a drastically different conformation from that found in pPGN. In hPGC, these residues run parallel to the active site flap (Leu73 to Gly76) that contains Tyr75. Tyr75 normally forms part of the active site in aspartic proteinases and makes a conserved hydrogen bond to the $N^{\varepsilon 1}$ atom of Trp39. The conformation of the Ser1 to Pro6 segment disrupts this interaction in hPGC (Figure 6.14). This is the major difference between the active sites of hPGC and pPGN. Other important active site residues such as Thr33, Ser35, Trp39 and Thr218 are in nearly identical conformations to what is seen in pPGN (Figure 6.15). There are some fairly drastic amino acid substitutions at the active site in the progastricsin family relative to the pepsinogens. Phe111, part of the S3 substrate binding pocket in pPGN is an asparagine residue in hPGC. Met245 in pPGN is a tyrosine or phenylalanine in hPGC. Gly188 in pPGN is a leucine in hPGC. There are other more subtle differences as well. There is some evidence to suggest that gastricsins have a different substrate specificity than the pepsins (N. Tarasova, unpublished data) and the above-mentioned amino-acid changes could encrypt some of these differences.

The activation process

How does progastricsin, upon a lowering of pH, develop an active site and then manage to hydrolyse one of its own peptide bonds? This is a fascinating problem, and our structural analysis of progastricsin gives a few clues as to what is happening. Previous work suggests the protonation of a single ionizable group with a pK_a near 2 (Glick *et al.*, 1989; Kageyama *et al.*, 1989) to be the rate determining step in the unimolecular conversion of pepsinogens and progastricsins. The most likely candidate is one of the active site aspartates, which once protonated, would make weaker electrostatic interactions with Lys37p. Lys37p is on a short but well-ordered 3_{10} helix of the prosegment. Amino acid sequences immediately before (residues 24p to 30p) and after (residues 39p to 2) this helix, however, are very disordered. This suggests that disrupting just a few interactions could dislodge the 3_{10} helix and the flanking regions from the active site. Tyr38p, which also interacts with Asp217, is already in an unfavorable conformation (χ_1 = -35°) and could easily undergo a χ_1 rotation to move out of the active site S1' pocket. Once this occurs, other portions of the propart could unfold and reposition themselves in the active site to be hydrolyzed. In human progastricsin, the initial hydrolysis site is between Phe26p and Leu27p, under unimolecular cleavage conditions (Foltmann & Jensen, 1982). This portion of the prosegment is disordered,

but is also at the C-terminus of the second α-helix. Interestingly, the intramolecular cleavage site in porcine pepsinogen (Leu16p to Ile17p) is near the C-terminus of the first α-helix of the propart. It is likely that helices undergoing denaturation would unwind from their ends, and hence the ends of these helices would be unwound and likely in an extended conformation in the acid denaturing conditions of a zymogen activating solution. The side chain of Tyr9, would also have to be displaced from the active site pocket for productive substrate binding to occur. This could happen concomitantly with the release of Lys37p from the active site, by a simple χ_1 rotation of the Tyr9 side chain. Alternatively, displacement of the 3_{10} helix containing Lys37p out of the active site could also cause the segment containing residues 9 to 14 of the mature enzyme to rearrange and move away from the active site. The protonation of Asp11, and weakening of the Arg14p-Asp11 salt bridge would assist such a conformational rearrangement.

REFERENCES

Abad-Zapatero, C., Rydel, T. J. & Erickson, J. (1990). Revised 2.3 Å structure of porcine pepsin: Evidence for a flexible subdomain. *Proteins: Struct., Funct. Genet.* **8**, 62-81.

Abad-Zapatero, C., Rydel, T. J., Neidhart, D. J., Luly, J. & Erickson, J. W. (1991). Inhibitor binding induces structural changes in porcine pepsin. In *Advances in Experimental Medicine and Biology*, vol. 306: Structure and Function of the Aspartic Proteinases (Dunn, B. M., ed.), pp. 9-21, Plenum, New York.

Al-Janabi, J., Hartsuck, J. A. & Tang, J. (1972). Kinetics and mechanism of pepsinogen activation. *J. Biol. Chem.* **247**, 4628-4632.

Andreeva, N. S. & James, M. N. G. J. (1991). Why does pepsin have a negative charge at very low pH? An analysis of conserved charged residues in aspartic proteinases. In *Advances in Experimental Medicine and Biology*, vol. 306: Structure and Function of the Aspartic Proteinases (Dunn, B. M., ed.), pp. 39-45, Plenum, New York.

Bank, R. A., Russell, R. B., Pals, G. & James, M. N. G. J. (1991). Consequences of intramolecular ionic interactions for the activation rate of human pepsinogens A and C as revealed by molecular modelling. In *Advances in Experimental Medicine and Biology*, vol. 306: Structure and Function of the Aspartic Proteinases (Dunn, B. M., ed.), pp. 101-105, Plenum, New York.

Baudys, M. & Kostka, V. (1983). Covalent structure of chicken pepsinogen. *Eur. J. Biochem.* **136**, 89-99.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer Jr., E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasmui, M. (1977). The Protein Data Bank: A computer-based archival file for macromolecular structures. *J. Mol. Biol.* **112**, 535-542.

Brünger, A. T., Kuriyan, J. & Karplus, M. (1987). Crystallographic R factor refinement by molecular dynamics. *Science* **235**, 458-460.

Bustin, M. & Conway-Jacobs, A. (1971). Intramolecular activation of porcine pepsinogen. *J. Biol. Chem.* **246**, 615-620.

Chen, L., Erickson, J. W., Rydel, T. J., Park, C. H., Neidhart, D., Luly, J. & Abad-Zapatero, C. (1992). Structure of a pepsin/renin inhibitor complex reveals a novel crystal packing induced by minor chemical alterations in the inhibitor. *Acta Cryst.* **B48**, 476-488.

Chothia, C. & Lesk, A. M. (1982). Evolution of proteins formed by β-sheets. I. Plastocyanin and Azurin. *J. Mol. Biol.* **160**, 309-323.

Cooper, J. B., Khan, G., Taylor, G., Tickle, I. J. & Blundell, T. L. (1990). X-ray analyses of aspartic proteinases II. Three dimensional structure of the hexagonal crystal form of porcine pepsin at 2.3 Å resolution. *J. Mol. Biol.* **214**, 199-222.

Foltmann, B. (1981). Gastric proteinases - structure, function, evolution and mechanism of action. *Essays in Biochem.* **17**, 52-84.

Foltmann, B. (1988a). Structure and function of prosegments in zymogens for aspartic proteinases. *Biol. Chem. Hoppe-Seyler* **369**, Suppl. 311-314.

Foltmann, B. (1988b). Activation of human pepsinogens. *FEBS Lett.* **241**, 69-72.

Foltmann, B. (1988c). *Aspartic proteinases: Alignment of amino acid sequences* In *Proceedings of the 18th Linderstrøm-Lang Conference.* (Foltmann, B., ed.), pp. 7-20, 4-8 July 1988, Elsinore, Denmark. University of Copenhagen, Copenhagen.

Foltmann, B. & Jensen, A. L. (1982). Human progastricsin: Analysis of intermediates during activation into gastricsin and determination of the amino acid sequence in the propart. *Eur. J. Biochem.* **128**, 63-70.

Fruton, J.S. (1976). The mechanism of the catalytic action of pepsin and related acid proteinases. *Adv. Enzymol.* **44**, 1-44.

Fujinaga, M. & Read, R. J. (1987). Experiences with a new translation-function program. *J. Appl. Cryst.* **20**, 517-521.

Glick, D. M., Shalitin, Y. & Hitt, C. R. (1989). Studies o[illegible]e irreversible step of pepsinogen activation. *Biochemistry* **28**, 2626-2630.

Harboe, M., Anderson, P. M., Foltmann, B., Kay, J. & Kassell, B. (1974).The activation of bovine pepsinogen. Sequence of peptides released, identification of a pepsin inhibitor. *J. Biol. Chem.* **249**, 4487-4494.

Harper, E. T. & Rose, G. D. (1993). Helix stop signals in proteins and peptides: the capping box. *Biochemistry* **32**, 7605-7609.

Hartsuck, J. A., Koelsch, G. & Remington, S. J. (1992). The high-resolution crystal and molecular structure of porcine pepsinogen. *Proteins* **13**, 1-25.

Hayano, T., Sogawa, K., Ichihara, Y., Fujii-Kuriyama, Y. & Takahashi, K. (1988). Primary structure of human pepsinogen C gene. *J. Biol. Chem.* **263**, 1382-1385.

Hendrickson, W. A. & Konnert, J. H. (1980). Incorporation of stereochemical information into crystallographic refinement. In *Computing in Crystallography* (Diamond, R., Ramaseshan, S. & Venkatesan, K., ed.), pp. 13.01-13.23, Indian Academy of Sciences, International Union of Crystallography, Bangalore, India,

Herriott, R. M. (1939). Kinetics of the formation of pepsin from swine pepsinogen and identification of an intermediate compound. *J. Gen. Physiol.* **22**, 65-78.

Hobart, P. M., Fugliano, M., O'Connor, B. A., Schaefer, I. M. & Chirgwin, J. M. (1984). Genomic DNA sequence analysis of human renin. *Proc. Natl. Acad. Sci. USA* **81**, 5026-5030.

Holm, I., Ollo, R., Panthier, J-J & Rougeon, F. (1984). Evolution of aspartyl proteases by gene duplication: the mouse renin gene is organised in two homologous clusters of four exons. *EMBO J.* **3**, 557-562.

Howard, A. J., Nielsen, C. & Xuong, Ng H. (1985). Software for a diffractometer with multiwire area detector. *Methods in Enzymology* **114**, 452-472.

Ichihara, Y., Sogawa, K., Morohashi, K., Fuji-Kuriyama, Y. & Takahashi, K. (1986). Nucleotide sequence of a nearly full-length cDNA coding for pepsinogen of rat gastric mucosa. *Eur. J. Biochem.* **161**, 7-12.

Ivanov, P. K., Chernaya, M. M., Gustchina, A. E., Pechik, I. V., Nikonov, S. V. & Tarasova, N. I. (1990). Isolation, crystallization and preliminary X-ray diffraction data of human progastricsin. *Biochim. Biophys. Acta Biochim. Biophys. Acta* **1040**, 308-310.

James, M. N. G. & Sielecki, A. R. (1986). Molecular structure of an aspartic proteinase zymogen, porcine pepsinogen, at 1.8 Å resolution. *Nature* **319**, 33-38.

Jones, T. A. (1978). A graphics model building and refinement system for macromolecules. *J. Appl. Cryst.* **11**, 268-272.

Kageyama, T. & Takahashi, K. (1986a). The complete amino acid sequence of monkey progastricsin. *J. Biol. Chem.* **261**, 4406-4419.

Kageyama, T. & Takahashi, K. (1986b). The complete amino acid sequence of monkey pepsinogen A. *J. Biol. Chem.* **261**, 4395-4405.

Kageyama, T., Ichinose, M., Miki, K., Athauda, S., Tanji, M. & Takahashi, K. (1989). Difference of activation processes and structure of activation peptides in human pepsinogens A and progastricsin. *J. Biochem. (Japan)* **105**, 15-22.

Kageyama, T., Ichinose, M., Tsukada, S., Miki, K., Kurokawa, K., Koiwai, O., Tanji, M., Yakabe, E., Athauda, S. B. P. & Takahashi, K. (1992). Gastric procathepsin E and progastricsin from guinea pig. *J. Biol. Chem.* **267**, 16450-16459.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283-291.

McPhie, P. (1972). A spectrophotometric investigation of the pepsinogen-pepsin conversion. *J. Biol. Chem.* **247**, 4277-4281.

Miller, M., Jaskólski, M., Rao, J. K. M., Leis, J. & Wlodawer, A. (1989). Crystal structure of a retro viral protease proves relationship to aspartic protease family. *Nature* **337**, 576-579.

Navaza, J. (1994). Amore: an automated package for molecular replacement. *Acta Cryst.* **A50**, 157-163.

Otwinowski, Z. (1990). Accurate refinement of heavy atom parameters. Amer. Crystallogr. Assoc. 1990 Annual Meeting, Abstr. C04.

Pals, G., Azuma, T., Mohandas, T. K., Bell, G. I., Bacon, J., Samloff, I. M., Walz, D. A., Barr, P. J. & Taggart, R. T. (1989). Human pepsinogen C (progastricsin) polymorphism: evidence for a single locus located at 6p21.1-pter. *Genomics* **4**, 137-145.

Pearl, L. H. & Taylor, W. R. (1987). A structural model for retroviral proteases. *Nature* **329**, 351-354.

Perlmann, G. E. (1963). The optical rotatory properties of pepsinogen. *J. Mol. Biol.* **6**, 452-464.

Read, R. J. (1986). Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Cryst.* **A42**, 140-149.

Read, R. J. & Schierbeek, A. (1988). A phased translation function. *J. Appl. Cryst.* **21**, 490-495.

Rees, D. C., Lewis, M. & Lipscomb, W. N. (1983). Refined crystal structure of carboxypeptidase A at 1.54 Å resolution. *J. Mol. Biol.* **168**, 367-387.

Richardson, J. S. & Richardson, D. C. (1988). Amino acid preferences for specific locations at the ends of α helices. *Science* **240**, 1648-1652.

Samloff, I. M. (1982). Pepsinogens I and II: Purification from gastric mucosa and radioimmunoassay in serum. *Gastroenterology* **82**, 26-33.

Schechter, I. & Berger, A. (1967). On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* **27**, 157-162.

Sielecki, A. R., Fedorov, A. A., Boodhoo, A., Andreeva, N. S. & James, M. N. G. (1990). Molecular and crystal structures of monoclinic porcine pepsin refined at 1.8 Å resolution. *J. Mol. Biol.* **214**, 143-170.

Sielecki, A. R., Fujinaga, M., Read, R. J. & James, M. N. G. (1991). Refined structure of porcine pepsinogen at 1.8 Å resolution. *J. Mol. Biol.* **219**, 671-692.

Sogawa, K., Fuji-Kuriyama, Y., Mizukami, Y., Ichihara, Y. & Takahashi, K. (1983). Primary structure of human pepsinogen gene. *J. Biol. Chem.* **258**, 5306-5311.

Taggart, R. T., Mohondas, T. K., Shows, T. B. & Bell, G. I. (1985). Variable numbers of pepsinogen genes are located in the centromeric region of human chromosome 11 and determine the high-frequency electrophoretic polymorphism. *Proc. Natl. Acad. Sci. USA* **82**, 6240-6244.

Taggart, R. T., Cass, L. G., Mohandas, T. K., Derby, P., Barr, P. J., Pals, G. & Bell, G-I (1989). Human pepsinogen C (progastricsin). Isolation of cDNA clones, localization to chromosome 6, and sequence homology with pepsinogen A. *J. Biol. Chem.* **264**, 375-379.

Tang, J., James, M. N. G., Hsu, I. N., Jenkins, J. A. & Blundell, T. L. (1978). Structural evidence for gene duplication in the evolution of the acid proteases. *Nature* **271**, 618-621.

ten Kate, R. W., Pals, G., Res, J. C., van Kamp, G. J., Verheugt, F. W., Pronk, J. C., Donker, A. J., Eriksson, A. W. & Mewwissen, S. G. (1989). The glomerular sieving of pepsinogens A and C in man. *Eur. J. Clin. Invest.* **19**, 306-310.

Thiessen, W. E. & Levy, H. A. (1973). International Union of Crystallography World List of Crystallographic Computer Programs. *J. Appl. Crystallogr.* **6**, 309-309.

Toh, H., Ono, M., Saigo, K. & Miyata, T. (1985). Retroviral protease-like sequence in the yeast transposon Ty1. *Nature* **315**, 619.

Tronrud, D. E. (1992). Conjugate-direction minimization. An improved method for the refinement of macromolecules. *Acta Cryst.* **A48**, 912-916.

Tronrud, D. E., TenEyck, L. F. & Matthews, B. W. (1987). An efficient general-purpose least-squares refinement program for macromolecular structures. *Acta Cryst.* **A43**, 489-503.

Xuong, N. H., Sullivan, D., Neilsen, C. & Hamlin, R. (1985). Use of the multiwire area detector diffractometer as a national resource for protein crystallography. *Acta Cryst.* **B41**, 267-269.

CHAPTER 7

DISCUSSION

NFP and bacterial bioluminescence

The three dimensional structure of the nonfluorescent flavoprotein (NFP) from *Photobacterium leiognathi* reveals the protein to have a fold that appears to be derived from a $(\beta/\alpha)_8$ barrel motif such as that present in triose phosphate isomerase (TIM) and glycolate oxidase. NFP only has seven β-strands and the barrel has a gap in it. The gap is between barrel strands β_2 and β_3 and is part of the extensive dimer interface found in NFP. NFP has evolved as a dimeric protein and, interestingly has a dimeric evolutionary cousin, bacterial luciferase. Unlike NFP, which is a homodimer, luciferase is a heterodimer of α and β subunits, each subunit encoded as a separate gene on the bacterial *lux* operon. The subunits of luciferase share about 30% of their amino acids in alignments and clearly have descended from a common ancestor. NFP is also a cousin of the luciferase subunits, sharing about 16% and 30% of its 228 amino acids in sequence alignments with the α and β subunits of luciferase, respectively. Curiously, both subunits of luciferase contain an additional region of 100 amino acids that is not found in NFP. We have shown that this region possibly encodes enough secondary structural units for the luciferase subunits to fold as $(\beta/\alpha)_8$ barrels. The luciferase α subunit likely has its active site at the barrel's C-terminus, just as in all other $(\beta/\alpha)_8$ barrel enzymes. However, this suggests that although NFP and luciferase have evolved as dimeric moieties, they have not chosen the same dimerization strategy. The hypothesis that the luciferase subunits fold as $(\beta/\alpha)_8$ barrels negates the dimer interface structure seen in NFP. The insertion required for the completion of the barrel in luciferase falls right at the NFP dimer interface.

NFP's closest structural relative in the current repertoire of $(\beta/\alpha)_8$ enzymes is glycolate oxidase (Chapter 4). Although differing in key details of the polypeptide chain fold, the packing of hydrophobic residues in the central core of the β barrel of NFP and glycolate oxidase are strikingly similar (Chapter 4). Helices that contribute to the β/α topology in both proteins have similar relative orientations. Strands β_1 and β_7 in NFP are structurally equivalent to strands β_1 and β_8 of glycolate oxidase. This means that in both proteins the first and last strands of the barrel are hydrogen bonded to each other, suggesting the presence, at one time, of a complete 8-stranded barrel in the common

ancestor of NFP and bacterial luciferase. All known $(\beta/\alpha)_8$ barrel proteins initiate the $(\beta/\alpha)_8$ topology with a β strand (Leboida & Stec, 1988).

The predicted $(\beta/\alpha)_8$ topology of the luciferase subunits suggests the presence of a small domain at the N-terminus of the β-barrel (Chapter 4). Such a domain is found in glycolate oxidase and other structurally related $(\beta/\alpha)_8$ enzymes and partially covers the N-terminal opening of the barrel (Farber & Petsko, 1990). In the luciferase subunits, this proposed domain would be between the fourth helix and the fifth strand of the barrel, whereas in glycolate oxidase and other flavin mononucleotide (FMN) binding $(\beta/\alpha)_8$ enzymes it occurs before the first strand of the β-barrel. Glycolate oxidase and bacterial luciferase both utilize FMN as a cofactor, although in mechanistically unrelated reactions. Our structural comparison of NFP and glycolate oxidase suggests it is feasible for the luciferase α subunit to bind the FMN ribityl phosphate moiety in a similar fashion to that observed in glycolate oxidase and other $(\beta/\alpha)_8$ enzymes that bind phosphate (Wilmanns *et al.*, 1991). In addition, several of these enzymes contain a flexible loop between the sixth strand and the sixth helix of the barrel that appears to be involved in phosphate binding. The α subunit of bacterial luciferase contains a highly-conserved proteinase-sensitive region that is protected from proteolysis in the presence of phosphate (Baldwin & Ziegler, 1992). This loop is 27 amino acids in length and is absent from the β subunit of luciferase. Secondary structure predictions indicate that this region is not likely to form α-helices or β-strands (Chapter 4). Our modelling based on the structure of NFP would put this loop between the seventh canonical strand and the seventh canonical helix of the proposed luciferase $(\beta/\alpha)_8$ barrel.

Another interesting observation is that many $(\beta/\alpha)_8$ barrel enzymes activate a hydrogen atom next to a carbonyl group (Leboida & Stec, 1988; Rose, 1982). Bacterial luciferase involves the direct oxidation of a carbonyl compound, myristic aldehyde. Could it be that luciferase evolved from an FMN-dependent fatty acid oxidizing enzyme such as a β-hydroxylase? β-hydroxybutyrate biosynthesis and other secondary metabolic pathways are known to be connected to bacterial bioluminescence through regulation by the autoinducer (Meighen, 1993).

We have put forward arguements that suggest the subunits of bacterial luciferase are $(\beta/\alpha)_8$ barrel proteins. Unfortunately this does not lend much insight to the biological role of NFP, or the origin of its strange myristylated flavin chromophore. A likely hypothesis is that NFP is the result of a relatively recent evolutionary adaptation.

We were suprised to find that neither the FMN nor myristate binding sites on NFP appear to be conserved in either of the luciferase subunits (Chapter 3). It seems odd that the luciferase α subunit has evolved to selectively bind FMN and myristic aldehyde, only to evolve into a protein (the β subunit) that confers stability on its predecessor (O'Kane & Prasher, 1992). Then the β subunit evolves into NFP and reacquires the ability to bind FMN and myristate.

It seems to be an intuitively roundabout way to obtain a necessary biological property. We would, however, like to propose a biological function for NFP. The function hinges on the hypothesis that the myristylated FMN's bound by NFP are themselves created at the luciferase active site in an enzymatic process. Our proposed function also requires that formation of the NFP chromophore be a relatively minor side reaction of the 4a-peroxy-4a,5-dihydroflavin with myristic aldehyde, and that the resultant adduct be a potent inhibitor of bacterial luciferase (via product inhibition). Over time, such compounds would be expected to build up to a level where a significant fraction of the cellular luciferase pool is inhibited by one of these molecules being bound at the active site on the α subunit. This is exactly what has been observed in cultures of bioluminescent bacteria. Luciferase activity decreases with increasing time in initially bright cultures of cells. Mutant strains of bacteria that have a defect in aldehyde biosynthesis do not exhibit this time-dependent inhibition (Hastings *et al.*, 1985). We suggest that the NFP chromophore is produced by a non-luminescent side reaction at the luciferase active site and that it binds tightly enough to the active site to interfere with the bioluminescence reaction. If the *Photobacterium* luciferases were particularly susceptible to a side reaction that compromised their bioluminescence output through formation of an inhibitory compound, the bacteria would require a means to remove the inhibitor or they would suffer from reduced light output. A product inhibitor should have chemical characteristics of substrates involved in the bioluminescence reaction. The inhibitor should also have a configuration close enough to that of the bound substrates to easily fit into the active site and inhibit the enzyme. A model 4a-peroxy-4a,5-dihydroflavin structure can be constructed that puts C3' of the myristic aldehyde in spatial proximity of the isoalloxazine C6 atom, and is entirely consistent with the NFP chromophore structure. The fact that only the S-enantiomer at C3' of the fatty acid is observed, strongly suggests the participation of an enzyme in the formation of this unusual molecule. However, the possibility exists that we are observing selective binding to NFP of only one of the two potential enantiomers of the adduct. We suspect that this

side reaction occurs more frequently in *Photobacterium* species as a result of differences at the active site relative to the *Vibrio* genus. *Photobacterium and Vibrio* luciferases are clearly distinct enzymes based on their respective amino acid sequences. Mutagenesis experiments have shown that alteration of a single residue in the vicinity of the *Vibrio harveyi* luciferase active site can drastically change the relative yields of the various possible reaction pathways (Xi *et al.*, 1990). If such an inhibitory reaction was operational in bioluminescent bacteria, one would expect natural selection to find a way to get around the problem to maximize light output. This of course assumes that increased light production is a selectable phenotypic trait and bacteria that produce more light have a selective advantage over their dimmer relatives. Bacteria could get around the problem of having to deal with such an inhibitor by just making more luciferase, but a more energy efficient strategy would be to produce a protein that binds the inhibitor more tightly than luciferase and with a more efficient stoichiometry. In fact NFP can bind six times as many myristlylated flavin chromophores as luciferase per number of amino acids synthesized in the respective proteins. We believe that the most likely biological function of NFP is to bind these myristylated flavins in the cell, and sequester them from luciferase, so that luciferase, in turn, can go about emitting light. This hypothesis could be tested in various ways. Firstly the NFP chromophore could be isolated from NFP and its effect on luciferase examined *in vitro*. Secondly, strains of *Photobacterium leiognathi* could be constructed that differ only in the presence or absence of NFP. Dark mutants of these bacteria,that have defects in aldehyde biosynthesis, could be examined to see if they produce the NFP chromophore in significant amounts.

Human progastricsin

The 1.62 Å refined structure of human progastricsin (hPGC) suggests that all digestive aspartic proteinases share a common mechanism of zymogen inhibition and acid pH conversion to the mature enzyme. Specifically Lys37p and Tyr38p play identical roles to those previously ascribed to Lys36p and Tyr37p in porcine pepsinogen (James & Sielecki, 1986; Sielecki *et al.*, 1991). Despite the multitude of positively charged residues in the prosegment, there is only one other electrostatic interaction common to the prosegments of the two zymogen structures. This is the salt bridge between Arg14p and Asp11. Asp11 occupies a portion of the active site cleft in the zymogen, and moves some distance away in the active enzyme. A drop in pH would weaken this conserved salt bridge and possibly bring about conformational changes that

trigger the activation process. The roles of these amino acid residues in zymogen activation could be tested by site specific mutagenesis experiments. ^{15}N NMR experiments also come to mind when thinking of possible molecular probes for insights into aspartic proteinase zymogen conversion.

In addition to the many similarities between human progastricsin and porcine pepsinogen, there are some striking differences in conformation between the two proteins. Of particular note is the unusual conformation of the active site flap in human progastricsin, in sharp contrast to most other aspartic proteinase structures. This movement of the flap is mostly due to the C-terminus of the propart (Arg38p-Pro6) being in drastically different conformations in the two zymogens. Is this a distinguishing feature of the progastricsins and the pepsinogens or is it simply a result of crystal packing forces? Subsequent studies of other aspartic proteinase zymogens should answer this question.

REFERENCES

Baldwin, T. O. & Ziegler, M. M. (1992). The biochemistry and molecular biology of bacterial bioluminescence. In *Chemistry and Biochemistry of Flavoenzymes* (Müller, F., ed.), vol. III, pp. 467-530, CRC, Boca Raton, FL.

Farber, G. K. & Petsko, G. A. (1990). The evolution of α/β barrel enzymes. *TIBS* **15**, 228-234.

Hastings, J. W., Potrikus, C. J., Gupta, S. C., Kurfürst, M. & Makemson, J. C. (1985). Biochemistry and physiology of bioluminescent bacteria. In *Advances in Microbial Physiology* (Rose, A. H. & Tempest, D. W., eds.), vol. 26, pp. 235-291, Academic Press, London/New York.

James, M. N. G. & Sielecki, A. R. (1986). Molecular structure of an aspartic proteinase zymogen, porcine pepsinogen, at 1.8 Å resolution. *Nature* **319**, 33-38.

Lebioda, L. & Stec, B. (1988). Crystal structure of enolase indicates that enolase and pyruvate kinase evolved from a common ancestor. *Nature* **333**, 683-686.

Meighen, E. A. (1993). Bacterial bioluminescence: organization, regulation, and application of the *lux* genes. *FASEB J.* **7**, 1016-1022.

O'Kane, D. J. & Prasher, D. C. (1992). Evolutionary origins of bacterial bioluminescence. *Mol. Microbiol.* **6**, 443-449.

Rose, I. A. (1982). Enzymology of enol intermediates. In *Methods in Enzymology* **87**, 84-97 (Purich, D. L., ed.), Academic Press, New York.

Sielecki, A. R., Fujinaga, M., Read, R. J. & James, M. N. G. (1991). Refined structure of porcine pepsinogen at 1.8 Å resolution. *J. Mol. Biol.* **219**, 671-692.

Wilmanns, M., Hyde, C. C., Davies, D. R., Kirschner, K. & Jansonius, J. N. (1991). Structural conservation in parallel β/α-barrel enzymes that catalyse three sequential reactions in the pathway of tryptophan biosynthesis. *Biochemistry* **30**, 9161-9169.

Xi, L., Cho, K.-W., Herndon, M. E. & Tu, S.-C. (1990). Elicitation of an oxidase activity in bacterial luciferase by site-directed mutation of a noncatalytic residue. *J. Biol. Chem.* **265**, 4200-4203.